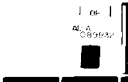
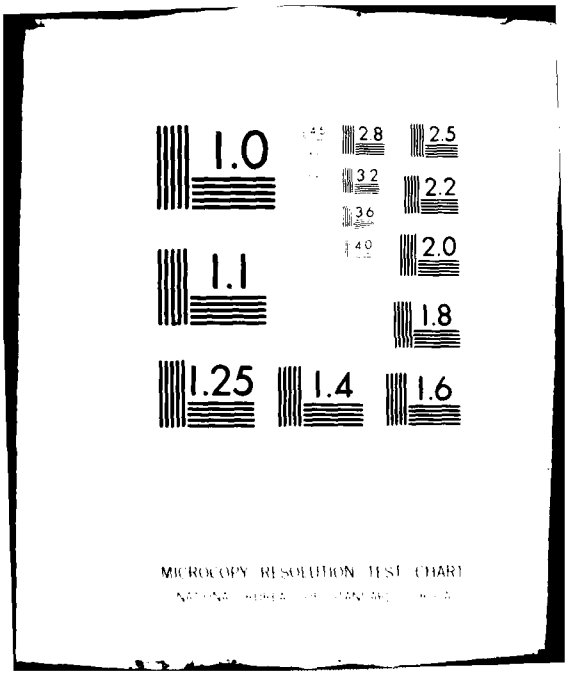
AD-A089 832 MASSACHUSETTS INST OF TECH CAMBRIDGE F/G 5/8
EFFICIENT COMPUTATIONS AND REPRESENTATIONS OF VISIBLE SURFACES.(U)
DEC 79 W RICHARDS, K A STEVENS AFOSR-79-0020

UNCLASSIFIED AFOSR-TR-80-0966 NL



END
DATE FILMED
11-80
DTIC


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART

(18) AFOSR (19) TR-80-0966

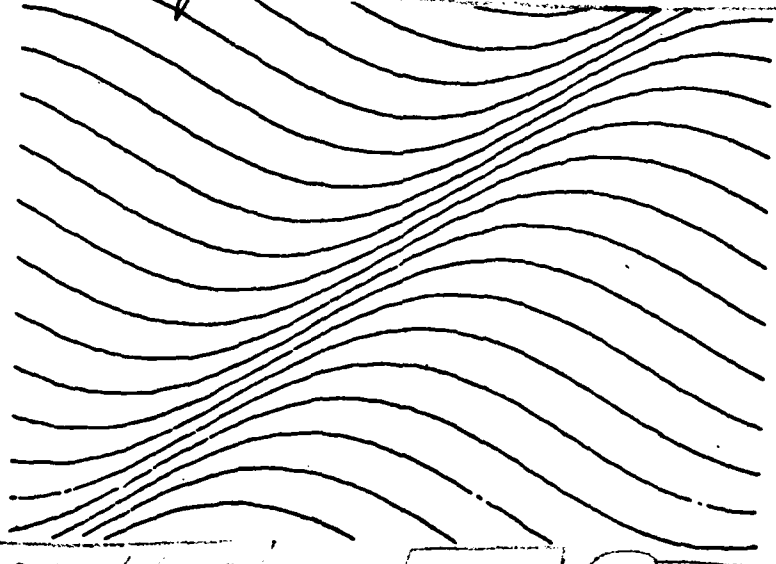
(6) EFFICIENT COMPUTATIONS AND REPRESENTATIONS OF VISIBLE SURFACES.

by

(10) WHITMAN RICHARDS
DEPARTMENT OF PSYCHOLOGY
and
KENT A. STEVENS
ARTIFICIAL INTELLIGENCE LABORATORY

LEVEL II

(9) Final rept. 1 Oct 78 - 31 Oct 79.



(15) AFOSR-79-0020 (16) 2313 (17) A2

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASS. 02139

(12) 75

(11) Dec[redacted]79

Approved for public release; distribution unlimited.

Final Report
Contract Number: AFOSR-79-0020
AFOSR Project Manager: Major J.A. Thorpe

Prepared for
Air Force Office of Scientific Research
Air Force Systems Command
Bolling Air Force Base, D.C. 20332

220000

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: AFOSR-TR- 80-0966	2. GOVT ACCESSION NO.: AD-A089832
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): EFFICIENT COMPUTATIONS AND REPRESENTATIONS OF VISUAL SURFACES	5. TYPE OF REPORT & PERIOD COVERED: FINAL REPORT 10/1/78 - 10/31/79
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Whitman Richards and Kent A. Stevens	8. CONTRACT OR GRANT NUMBER(s): AFOSR-79-0020 new
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Massachusetts Institute of Technology / 77 Massachusetts Avenue Cambridge, Mass. 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2313/A2
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research (NL) Bolling Air Force Base, Washington, D.C.20332	12. REPORT DATE: December, 1979
	13. NUMBER OF PAGES: 73
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Vision, surface contours, transparency

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

When the visual system is approached as performing information processing tasks, the processor can be characterized as solving a computational problem (Marr, 1976; Marr & Poggio, 1977). The first step in this approach is to formulate the problem precisely. This report addresses two such problems, the extraction of surface shape information from the image of contours lying on a surface and the determination of transparency. Our preliminary analyses discuss

--continued--

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

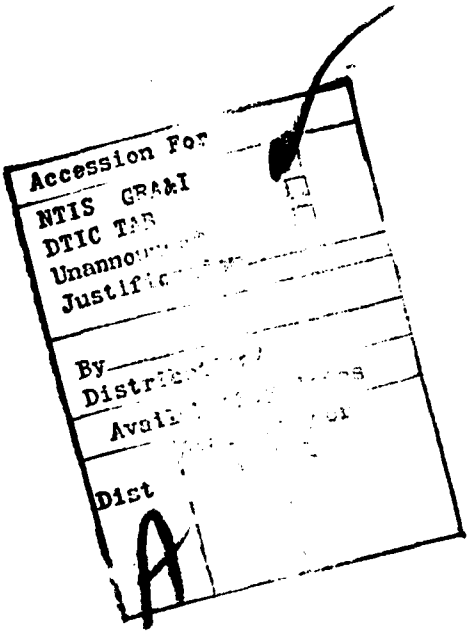
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

the constraints and assumptions that underly the perception of surface shape and transparency. The identification and elucidation of these assumptions and constraints is important for the design of flight simulators, in order that synthetic images are not built up from conflicting ones that will mislead the observer.

The first part of the report examines our perception of surface contours (e.g., the projections of linear surface markings such as seams, wrinkles, pigmentation edges, glossy reflections, and shadow edges). Generally such contours lie interior to the silhouette of an object and have been regarded as merely contributing to texture, or to making the surface appear solid, or simply adding to the complexity of the image. In fact, they constitute an important source of information about surface shape. This study addresses a few fundamental issues concerning our perception of shape from contour: the information content of surface contours (remembering that we usually have little *a priori* knowledge that we may bring to bear in their interpretation), the computational problems of inferring shape from contour, and the sorts of constraints that would be sufficient for solving these problems. The approach taken is to decompose the problem into two subproblems: (1) relating the surface contour in the image to the corresponding curve in 3-D (called the *contour generator*) and (2) relating the contour generator to the actual surface on which it lies. This decomposition allows one to focus on the aspect of the problem governed by projective geometry (the relation between the contour generator and the surface).

The second part of this report examines the problem of determining transparency from visual information. Transparency of a homogeneous surface is considered as a limiting case of a screen consisting of a net of opaque elements. When the elements of the screen become invisible, then the occluded scene is viewed through a transparent film. The basic problem of transparency is to determine which part of the content of an image is attributable to the screen, and which part arises from the occluded scene behind the screen. Although spatial (and temporal) as well as the spectral and intensity content of an image may be used for such judgements, this study focuses upon the color and lightness information available to the observer. The most useful cue for assessing transparency using these two stimulus dimensions is a reduction in contrast or chromatic content. This reduction can be determined locally from image information alone. Although such a solution is not computational optimal, it may be very practical for most natural scenes.

Contents



EFFICIENT COMPUTATIONS AND REPRESENTATIONS OF VISIBLE SURFACES

ABSTRACT

When the visual system is approached as performing information processing tasks, the processor can be characterized as solving a computational problem [Marr, 1976; Marr & Poggio, 1977]. The first step in this approach is to formulate the given problem precisely. This report addresses two such problems, the extraction of surface shape information from the image of contours lying on a surface and the determination of transparency. Our preliminary analyses discuss the constraints and assumptions that underly the perception of surface shape and transparency. The identification and elucidation of these assumptions and constraints is important for the design of flight simulators, in order that synthetic images are not built up from conflicting ones that will mislead the observer.

The first part of the report examines our perception of surface contours (e.g., the projections of linear surface markings such as seams, wrinkles, pigmentation edges, glossy reflections, and shadow edges). Generally such contours lie interior to the silhouette of an object and have been regarded as merely contributing to texture, or to making the surface appear solid, or simply adding to the complexity of the image. In fact, they constitute an important source of information about surface shape. This study addresses a few fundamental issues concerning our perception of shape from contour: the information content of surface contours (remembering that we usually have little *a priori* knowledge that we may bring to bear in their interpretation), the computational problems of inferring shape from contour, and the sorts of constraints that would be sufficient for solving these problems. The approach taken is to decompose the problem into two subproblems: (1) relating the surface contour in the image to the corresponding curve in 3-D (called the *contour generator*) and (2) relating the contour generator to the actual surface on which it lies. This decomposition allows one to focus on the aspect of the problem governed by projective geometry (the relation between the image surface contour and its contour generator) independently from that governed by intrinsic geometry (the relation between the contour generator and the surface).

The second part of this report examines the problem of determining transparency from visual information. Transparency of a homogeneous surface is considered as a limiting case of a screen consisting of a net of opaque elements. When the elements of the screen become invisible, then the occluded scene is viewed through a transparent film. The basic problem of transparency is to

determine which part of the content of an image is attributable to the screen, and which part arises from the occluded scene behind the screen. Although spatial (and temporal) as well as the spectral and intensity content of an image may be used for such judgements, this study focuses upon the color and lightness information available to the observer. The most useful cue for assessing transparency using these two stimulus dimensions is a reduction in contrast or chromatic content. This reduction can be determined locally from image information alone. Although such a solution is not computationally optimal, it may be very practical for most natural scenes.

PART I: SURFACE CONTOUR ANALYSIS

Kent A. Stevens

1. INTRODUCTION

This part describes geometrical constraints that may govern the way in which we perceive surface shape from surface contours in an image. In figure 1, for example, the smooth curves are seen in 3-D as lying on an undulating surface. We appreciate not only the shape of the surface, but also its spatial orientation relative to us, and to some extent we perceive the overall surface as receding in depth. The difficulty we face in interpreting figure 1 as merely a two-dimensional family of sinusoids (which it is) shows that we impose constraints in the form of *a priori* assumptions. Some of these assumptions lead us to interpret certain curves in the image as being *surface contours* (which correspond to actual curves across 3-D surfaces); others constrain the inferred surface shape that we derive by analysis of the surface contours. For the surface percept to be both definite and accurate, such constraints must define a unique surface, and must generally be valid.

Although many have considered our perception of the shape of contours (e.g., [Koffka, 1935, Rock, 1975]), the problem of how *surface shape* can be correctly inferred from the observed surface contours has received virtually no attention. The primary intentions of this part of the report are

(a) to formalize the computational problem of recovering 3D surface shape from contour,

(b) to introduce useful and valid constraints towards its solution, and

(c) to describe why those constraints are useful.

1.1 What information is carried by surface contours?

The contours in figure 1 are in orthographic[1] projection; hence we cannot derive distance

1. Orthographic projection is equivalent to a parallel projection, as opposed to a perspective projection. Figure 1 demonstrates that we may perceive shape from surface contours in orthographic projection. Later we will see that assuming that the projection is orthographic (and not perspective from some unknown viewing geometry) is probably necessary in the analysis.

information from perspectivity in the image. But the shape of the contours does provide surface shape information in two forms. In the vicinity of the surface contour one may deduce either:

> *surface orientation.* The relative surface orientation may be solved uniquely (i.e., up to a slant reflection since the projection is orthographic) or only to within a restricted range of slant and tilt.
>
> *qualitative surface shape.* The intrinsic geometry of the surface may be deduced from the shape of the surface contours. The primitive descriptors might include "flat", "singly curved", "cylindrical", "doubly curved" and so forth. This sort of shape information is independent of the viewpoint.

This is not to say that a depth map may not be computed from the image, but that the geometry of contours in an orthographic image more directly constrains surface orientation and intrinsic geometry than distance -- the computation of a depth map would effectively require the intermediate computation of surface orientation.

Note that information about intrinsic surface shape serves two useful purposes: (a) it constitutes a primitive, coordinate-free shape descriptor, and (b) it constrains the values in any representation of surface orientation or distance. Suppose that it can be determined from the image that a surface region must be singly curved, then this restriction can be imposed on any independently computed distance or surface orientation representation -- the distance or surface orientation must vary in a manner consistent with a singly curved surface. Later we shall see the contribution of this qualitative shape constraint on the computation of "shape from shading" (c.f., [Horn, 1975]).

1.2 Contours and contour generators

It is valuable to distinguish between a *contour* in an image and the corresponding curve in 3-D, called the *contour generator*, that projects to that contour (see [Marr, 1976b]). The contour generator is a physical curve which lies across a surface, such as a boundary between patches of differing reflectance (e.g., a pigmentation marking), a discontinuity in illumination (e.g., a shadow edge cast across the surface) or a discontinuity in surface orientation (e.g., a crease). The contour generator may also correspond to the boundary of the surface from the given viewpoint.

So on the one hand, we have the contours in the image; on the other hand, their corresponding physical curves in 3-D, the contour generators. To make 3-D interpretations from the image contours we often need to understand what causes them -- whether they correspond to object boundaries, shadow edges, or what.

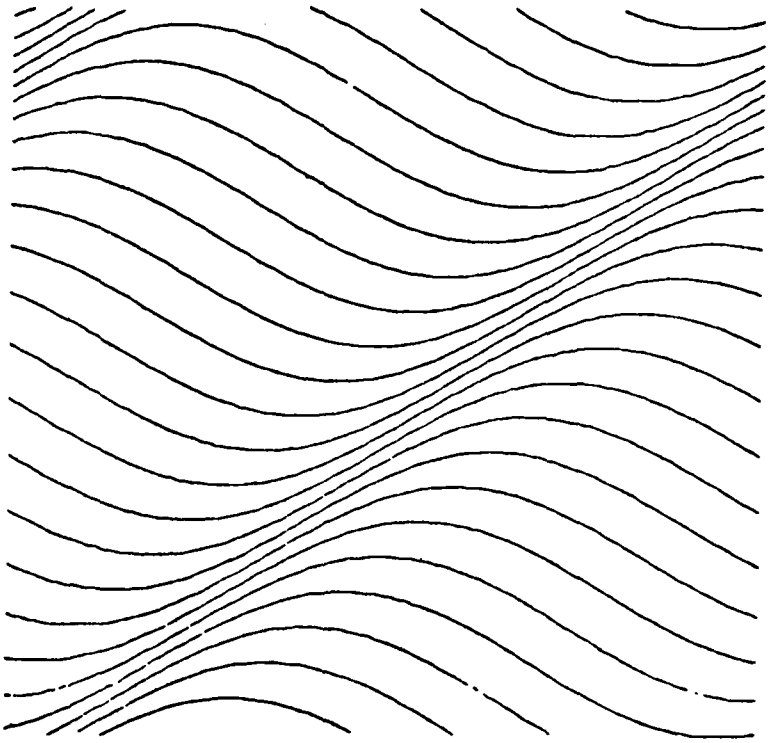


Figure 1. The undulating surface is suggested by a family of sinusoids. (This figure is adapted from Bridget Riley's *Kataraki 3*.) The curves are naturally interpreted as surface contours, i.e., the images of markings on a physical surface. What constraints can be brought to bear in making this 3-D interpretation?

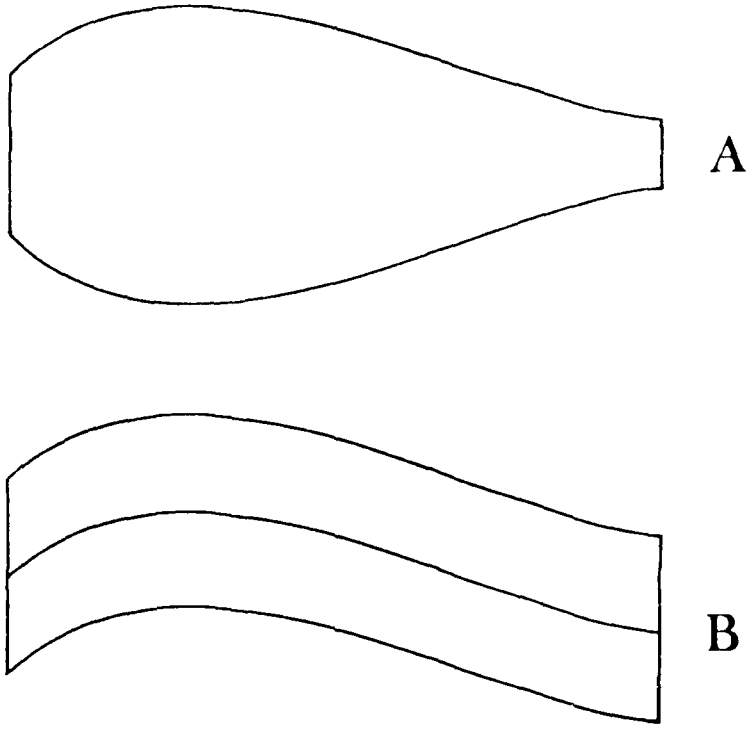


Figure 2. The curves in *a* are interpreted as tangential contours and the underlying surface is seen as a generalized cone, in this case, a vase-like 3D shape. Those in *b* are interpreted as surface contours and the surface appears like a gently curved flag or a ruled sheet of paper.

One basic distinction that is often proposed is between *object outlines* (also termed *bounding contours* or *occluding contours*) which correspond to the edge of an object's silhouette from the given viewpoint, and those contours that lie internal to the silhouette (which Gibson has termed "inlines"). A slight variant would be to distinguish only those bounding contours that correspond to the silhouettes of *smooth objects*. This distinction is probably fundamental for reasons that will be given in the following.

1.3 Tangential contours and surface contours

Physical objects are often smooth, and their silhouettes alone provide a strong source of information about the overall shape [Marr, 1976b]. For instance, consider a vase. Its silhouette projected onto the retinal image might appear like the upper outline shown in figure 2*a*. The contour generator in 3-D of this outline is the locus of points along which the line of sight just grazes the surface of the vase. In this case, the contour that comprises the outline will be termed a *tangential contour*. An important class of such contours are those that exhibit qualitative symmetry across an axis, like the "vase" in Fig. 2*a*. If it is assumed that the corresponding surface is smooth then its silouette is that of a generalized cone whose 3-D shape is recoverable (given some other restrictions, see [Marr, 1976b]). In this case, the silhouette boundary is comprised of tangential contours. Note that the surface orientation (its slant and tilt) is known at all points along a tangential contour: the slant is $\pi/2$ and the tilt is perpendicular to the contour.

In the previous discussion the 3D object was assumed to have a smooth surface. However, for objects with angular faces (such as many man-made objects), or objects that are basically 2-D surfaces (e.g., a leaf), the surface orientation will be discontinuous along the contour generator which corresponds to the outline. Since the line of sight does not graze the surface along the edge, the silhouette boundary is not a tangential contour. Observe that the contours in figure 2*b*, which we interpret as the outline of a gently curved sheet, present a fundamentally different problem than the contours in figure 2*a*. Neither do we assume that the surface is smooth nor that the contours are tangential contours.

The distinction that I propose is therefore not between "outlines" and "inlines" -- not whether the contour is along the boundary of the silhouette or interior to the bounary. Instead, the distinction is between the special case of outline contours, the tangential contours, and all other contours. This means that the outlines of objects that are not smooth will be treated as surface contours for our purposes. The reason for this is the following. The fact that a given contour is

part of an object outline does not constrain the shape of the underlying surface, expect when the surface is smooth. Otherwise, the contours merely delimit the visual extent of a object from the given viewpoint. The rest of this section will address the problem of using surface contours. In general, it will not concern us whether the surface contour is an outline contour as well.

1.4 Surface contours: structural and illumination

Thus far, we have only distinguished between contours corresponding to the silhouettes of smooth objects, *the tangential contours*, and all *other contours* (*those being collectively termed* surface contours). But there are various, distinct physical causes of these surface contours. In particular, we can distinguish two broad categories of surface contours, roughly speaking by whether the associated contour generator corresponds to a physical feature on the surface or merely due to illumination. The first category will be termed *structural contours*, the latter, *illumination contours*.

Structural contours are the projections of contour generators which mark some discontinuity on the surface, e.g. of reflectance or of surface orientation. Examples that occur in nature are given by the images of pigmentation markings on a zebra, wrinkles on skin, parallel ridges on leaves, rings on bamboo stalks, and cracks on wood or rock. Images of synthetic objects commonly present structural contours corresponding to seams, sharp edges, groves, and pigmentation markings.

Illumination contours are of three types: (a) the projections of glossy reflections, such as those that appear on metallic or wet surfaces, (b) the projections of shadow edges that have been cast upon a surface, and (c) the images of self-shadows, or "terminators" on surfaces. These three types have been grouped together as illumination contours because their presence is strongly dependent on the particular illumination and may shift their position relative to the surface as the viewpoint or light source geometry changes. They are all potentially useful sources of information about the shape of the surface, as we shall see, but since they depend on particular arrangements of illumination and viewing geometry, they may be considered as fortuitous.

We now turn to address the problem of constraining the visual interpretation of surface contours. First, we will examine a decomposition of the problem into two steps, each of which must be constrained. Constraints for each step are then introduced, and their validity discussed. Discussion of how these constraints are computationally useful is given in section 4.

2. THE CONSTRAINTS

In the following discussion a surface will be denoted by Σ, a contour generator by Γ, and the projection of Γ from viewpoint V will be the contour C_V (see figure 3). (When the viewpoint is not discussed, the contour will be referred to simply as C.)

A surface contour in the image is the projection[1] of a contour generator Γ lying on a surface Σ; neither the shape of Γ nor Σ is known *a priori*. Note that the surface contour C is completely determined by the 3-D locus of its generator Γ in space relative to the viewer, regardless of the orientation of the surface on which Γ lies so long as the surface allows Γ to be continuously visible along its length. This is an important point. We want to infer the shape of the surface Σ from the shape of the surface contour C, but in fact C is not a function of the shape Σ; C is only a function of Γ. In order to infer the shape of Σ, the relationship between Γ and Σ must be constrained. Likewise, to infer Γ from C, the relationship between Γ and C must be constrained. The decomposition that is suggested, therefore, involves two stages:

(a) inferring the shape of the contour generator in 3-space ($C \Rightarrow \Gamma$) then

(b) determining how the surface lies under the contour generator ($\Gamma \Rightarrow \Sigma$).

This can be thought of as (a) bending a wire in 3-space so that it appears to the viewer as does the contour in the image, then (b) gluing a ribbon along the wire to represent the strip of surface that lies directly under the contour generator. In these terms, we see that infinitely many bendings are possible that would appear identical from the given viewpoint, and the ribbon may twist arbitrarily along the wire. These two aspects of the problem are distinct.

This characterization applies equally to the problem of inferring surface shape from multiple surface contours $\{C_i\}$ in the image, such as those in figure 1. The geometrical arrangement of $\{C_i\}$, particularly if they are parallel, may constrain both stages I and II (section 4.2.2). Note that the appearance of figure 1 may lead one to suspect that parallelism uniquely constrains the surface, but the image is in orthographic projection and significantly different surfaces may project to the same

---

1. The projection is assumed orthographic, i.e., the contour generator Γ is assumed small compared to its viewing distance. The perspective distortions otherwise induced in C would be infeasible to differentiate from those induced by slight distortions in Γ under conditions of orthographic projection. Note further that the informal term "image plane" will be used, although the retinal projection is more closely approximated by spherical projection.

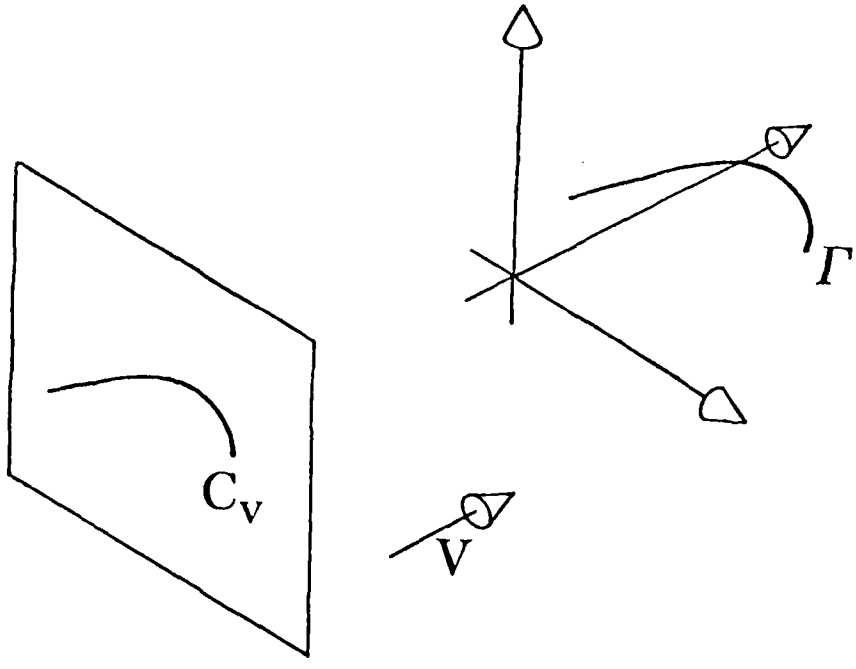


Figure 3. The orthographic projection of *contour generator* Γ from viewpoint V is C_v. The curve C_v is termed an *occluding contour* if it is an edge of the silhouette of an object from viewpoint V. In particular, if the line of sight just grazes the surface along Γ then the curve C_v is also a *tangential contour*. The curve C_v is termed a *surface contour* if it is not a tangential contour.

image -- the separation in depth between the contour generators on the surface is not restricted.[1] Thus even in the case of multiple parallel contours, the surface interpretation process must be constrained, and that constraint is naturally described in terms of the above two stages.

This decomposition provides a framework for applying constraints to the problem of inferring Σ from C. The constraints necessary for stage I involve projective geometry, for the problem is naturally one of "deprojecting" from the image curve to the curve in space. The constraints necessary for stage II do not involve projective geometry -- they do not depend on the particular viewpoint. Rather they involve intrinsic geometry, specifically the relationship between the curve on the surface and the surface itself.

2.1 Some geometrical concepts

This section reviews some concepts that are necessary for discussing the relation between a curve on a surface and the underlying surface itself. I shall review the notions of Gaussian curvature, lines of curvature, developable surfaces and cylinders, asymptotic curves, and geodesics (c.f. [Hilbert & Cohn-Vossen, 1952]).

To introduce Gaussian curvature, consider the family of normal sections at some point of a smooth surface, i.e., the contours that result from sections that contain the surface normal at that point. The various section contours through that point usually vary in curvature, with greatest and least curvature occurring at two principal directions. For example, contours parallel to the axis of a cylinder will be lines of least curvature (they are straight), whereas circles in planes perpendicular to the cylinder's axis will be lines of greatest Gaussian curvature. (For a sphere, the curvature is constant everywhere and hence this shape has no unique principal direction of curvature.)An important property of the two principal directions is that they are mutually orthogonal at every point on the smooth surface.

The *Gaussian curvature* at a point is the product of the greatest and least curvatures. The Gaussian curvature may be positive, negative, or zero, and for an arbitrary surface may vary continuously across the surface. For example, the curvature is positive on a smooth pebble, negative on a saddle surface, and zero on a cylinder (defined momentarily).

---

1. In fact, one consistent surface solution is given immediately by the sheet of paper on which figure 1 of this article is printed -- the parallel contour generators would be the ink on the page.

A *line of greatest* (or *least*) *curvature* is a curve whose tangent everywhere coincides with one of the two principal directions. Important examples are the cross sections and meridians of surfaces of revolution (which of these is the line of greatest curvature depends on the surface shape).

A *developable surface* is a surface with zero Gaussian curvature everywhere (i.e., the curvature in at least one of the principal directions vanishes). Thus the lines of least curvature are straight lines on a developable surface. Examples of developable surfaces are planes, cylinders, and helicoids. Informally, they correspond to the class of surfaces that may be made by twisting and curling a sheet of paper.

A *cylinder* is a developable surface where the lines of least curvature are parallel. Cylinders may be formed by curling a sheet without torsion -- it may be rolled into a tube or be rippled like a hanging curtain. It is useful to think of a cylinder as a one-dimensional surface.

An *asymptotic curve* is a locus of points on the surface where the Gaussian curvature is zero. By definition, all curves on developable surfaces are asymptotic. On the other hand, surfaces with everywhere positive Gaussian curvature (such as a sphere) have no asymptotic curves. And surfaces of negative Gaussian curvature must have asymptotic curves, since the principle curvatures are of opposite sign and for some direction between the principle directions at each point on the surface the curvature must vanish.

Finally, a *geodesic*, usually defined as the shortest path between two points on a surface, is also a curve whose principal normal[1] everywhere coincides with the surface normal. Importantly, the lines of greatest and least curvature on a cylinder are geodesics.

2.2 What constraints might be useful?

We now introduce some constraints that allow solutions to steps I and II. They are provided by restricting the geometrical properties of the contour generators, and restricting the relationship between the contour generators and the surface on which they lie. This section only tabulates the various geometric restrictions. Next, in section 3 we will discuss the validity of assuming that these restrictions hold in natural situations involving actual contour generators on physical surfaces and,

---

1. The principal normal to a planar curve is the perpendicular to the tangent to the curve and lies in the plane of the curve. The principal normal to a curve with torsion, similarly, is perpendicular to the tangent but lies in the osculating plane of the curve at that point (where the osculating plane is defined by two successive tangents at the given point). Note that we will often restict curves to be planar, so fixing the plane of a geodesic immediately fixes the normal to the surface.

in section 4, we will describe how the restrictions constrain the shape-from-contour analysis.

2.2.1 Constraints on the contour generator

With regard to step I, the 3-D shape of a contour generator Γ (corresponding to a given surface contour C) may be recovered if restrictions are imposed on Γ and on the viewing position. Some of these restrictions are listed below.

(a) *general position*, the viewpoint is not misleading. This allows one to infer properties of the contour generator Γ on the basis of the properties of its image, the surface contour C. For instance, if C is smooth then Γ is smooth; if $\{C_i\}$ are parallel then $\{\Gamma_i\}$ are parallel.

(b) *planarity*, Γ is planar. This reduces the problem of determining Γ to that of determining the orientation of the plane Π containing Γ. The plane Π is constrained by the following.

(c) *symmetry*. Given planarity and general position, if C presents evidence of symmetry then Γ is symmetric, and the orientation of Π must be consistent with Γ being symmetric.

(d) *minimum curvature variation*. Given planarity and general position, if the curvature of Γ is roughly constant then the variations in curvature apparent in C may be attributed to foreshortening. Consequently that plane Π that minimizes the variation in curvature of Γ would solve Γ.

2.2.2 Constraints on the relation between contour generator and surface

Given the contour generator Γ, the surface Σ may be solved if the relationship between Γ and Σ is restricted. If Γ is planar and lies on some plane Π then the relationship between the contour generator and the surface is naturally described in terms of the angle between Π and the tangent plane to Σ for points along Γ. The relation between the surface and the contour generator is quite simple if we make the strong restriction that this angle is constant along the length of Γ. That is to say, the plane containing the contour generator meets the surface at a constant angle. The two cases we will consider is when the angle is $\pi/2$ and zero.

If the angle between Π and the tangent plane to Σ is $\pi/2$, then:

Γ is *geodesic*. The surface normal coincides with the principal normal to Γ for points along Γ.

If the angle between Π and the tangent plane to Σ is zero, then:

> Γ is *asymptotic*. The surface normal coincides with the normal to Π for points along Γ, and furthermore, the Gaussian curvature of Σ for points along Γ is zero.

These two solutions, geodesic and asymptotic, form the basis for constraining the relation between the contour generator and the surface. Given general position and planarity, we also have an important restriction on Σ in the case of parallel surface contours $\{C_i\}$:

> $\{\Gamma_i\}$ are *parallel lines of curvature* and Σ is a *cylinder*. Furthermore, if the contour generators are geodesics, they are lines of greatest curvature; if asymptotics, the surface degenerates to be planar.

And finally, a derivative of the cylinder restriction may apply in the case of a single surface contour, if the corresponding contour generator is a line of greatest curvature and the surface is cylindrical, by the following restriction:

> Σ is *opaque*. The image of an individual line of greatest curvature on a cylinder allows some restriction on the shape of the surface.

Surface contours are often weak sources of information about the surface shape when analyzed individually, primarily because it is difficult to deduce the shape of the contour generators on an individual basis. The more important case probably involves the geodesic restriction on a collection of parallel contours taken together. Then the parallelism may be used to advantage in constraining the shape of both the contour generators and the surface on which they lie. Before pursuing the utility of these constraints any further, it is important to gain some insight into their validity.

3. WHEN ARE THE CONSTRAINTS VALID?

Do the contour generators in the real world meet these restrictions? In some situations it is valid to assume that a contour generator is, say, planar and geodesic, as we shall see. But there are also instances where the same assumptions are not valid -- the real world does not necessarily constrain the curves on surfaces to comply with any of the various ideal geometries. How often are the restrictions met in actuality? This is the issue of "ecological validity" discussed by Gibson, Brunswick, and others (c.f. [Gibson, 1950; Postman & Tolman, 1959]). We start with considering the validity of assuming general position.

3.1 General position

General position implies that the viewpoint is representative -- that the image taken from this position does not mislead us by accidental alignments. Two examples of viewpoints that are not general position may be imagined for a cube: In one instance the cube is positioned so that its silhouette is a regular hexigon. Equally misleading would be a cube positioned so that its silhouette is a perfect square.

When the assumption of general position is correct we may make valid deductions, in particular, deductions about contour generators. Two examples of these deductions which we shall pursue are the following: If a surface contour is smooth, the corresponding contour generator is smooth, and if surface contours are parallel, their contour generators are also parallel.

The contour generator need not be smooth simply because its projection is smooth: a discontinuity in tangent along a contour generator might be hidden from the given viewpoint -- the plane containing the discontinuity might also contain the line of sight so that the discontinuity would not be apparent. But if the distribution of spatial orientations of planes relative to the viewer is uniform, the likelihood of such an accidental alignment would be insignificant. Similarly, some non-parallel curves may be constructed such that they appear parallel from certain viewpoints, but the probability of achieving a viewing position that allows this alignment becomes insignificant as the curves diverge from parallelism in 3-space. [1]

---

1. Implicit in the above argument is the reasonable expectation that the instances of actual parallelism, straightness, and so forth, are more probable than accidental alignments.

3.2 Geometrical properties of structural contours

In general, the geometry of structural contours is not strongly constrained because the processes that cause them are varied and often random. There are, however, some types of physical markings that are well constrained.

The clearest examples, perhaps, involve synthetic objects. With reference to the objects about you, observe that the smooth surfaces of man-made objects are usually comprised of either (a) planar surfaces, (b) singly curved surfaces, in particular cylinders, or (c) surfaces of revolution. In general, the boundaries between surfaces are planar, primarily for reasons of fabrication. Again, because of convenience in manufacturing as well as utility, curved surfaces are usually sliced by normal sections. Thus joints between surfaces of an object comprise geodesics on one or the other of the joining surfaces. The end of a "tin can" would be an example. Surface markings other than seams or joints are often geodesics as well, particular when the markings are on cylinders. When the markings are also planar, they additionally constitute lines of curvature. This combination of properties, planarity and geodesic, is particularly common.

Markings on surfaces of revolution usually follow either the axis or some cross section. Hence these seams, edges, ridges, and pigmentation markings are lines of curvature, geodesic, and planar. (A notable exception can be found in the spiral seams on cardboard tubes. They are geodesic but nonplanar.)

Flexible surfaces, both natural and synthetic, tend to be noncompressible hence developable, and are therefore cylinders when not subjected to torsion. Wrinkles produced by compression tend to be lines of curvature.

Many biological forms may be approximated as being composed of generalized cones [Marr, 1976b]. These surfaces often have markings that follow cross sections and meridians on the surface, and therefore are also lines of curvature, geodesic, and planar. Biological objects are often bilaterally symmetric, such as leaves. Their axes of symmetry are often evidenced by physical markings, and symmetric patterns are usually arranged across that axis. The symmetry may be used to advantage to restrict the possible orientations that would be consistent with the 3-D form being symmetric.

3.3 Geometrical properties of illumination contours

3.3.1 Cast shadows

The edge of a shadow cast across a surface is a fortuitous source of information about surface shape. We are familiar with the effectiveness of the shadow a fence post cast upon snow in indicating the undulations in the surface. But to accurately analyze the surface from the image of the cast shadow, a number of variables must be known. There are essentially two projections involved: the projection of the shadow onto the surface (the edge of which becomes the contour generator Γ) and the subsequent projection of Γ onto the image plane (as contour C). Thus the contour C in the image depends on (a) the shape of the physical shadow-casting edge, (b) the position of the light source -- together they specify the bundle of rays that will be cast upon the surface -- and (c) the position of the shadow-casting edge relative to the surface, and finally (d) the shape of the surface itself.

To appreciate the complexity of shadow interpretation in the general case, consider again the image of a tree trunk shadow cast on snow. Suppose there is a kink along the shadow edge. Is that due to a sharp depression in the snow (for instance, is the shadow falling across a footprint) or is it due to a kink in the tree (and the snow itself is flat)? If analyzing the shape of the surface is attempted prior to knowing the above factors, some assumptions are necessary. In the approach suggested here, the assumptions are two:

the contour generator is *planar* and *geodesic*.

In terms of this example, the above translate into assuming the edge casting the shadow is straight and that its profile (determined by the sun position and the trunk) intersects the ground at a right angle. Then if there is an apparent kink in the shadow edge it will be attributed to the surface, not to the tree. (Incidentally, it is informative to observe the shadow cast on the flat ground by a young tree which has a crooked trunk. The ground often appears to undulate according to the curves in the cast shadow.)

So we should discuss how the planarity and geodesic restrictions help the shape analysis. First note that if the shadow-casting edge is straight the contour generator (the shadow edge cast across the surface) constitutes a planar section of that surface. That is, the contour generator lies in the plane defined by the straight shadow-casting edge and the point light source. In this case, we may already determine qualitative information about the surface shape. Given general position, if the

contour in the image corresponding to the shadow edge is straight, the surface is flat; if it is curved, the surface is curved. To determine more quantitative shape information requires that (a) the relation between the contour generator Γ and the surface be known, and (b) the orientation of the plane of Γ be known. Hence we introduce the geodesic assumption. That is to say, the shadow edge across the surface is assumed to be a normal section of the surface. Weak justification for this assumption derives from considering shadows cast on the ground: Since shadow-casting edges are usually vertical (e.g., tree trunks, building edges, telephone poles, fences), the edge of the shadow amounts to a normal section, i.e., the shadow edge is roughly geodesic.

When do multiple, parallel sections occur in real situations? We may disregard the shadow of a picket fence as being artificial, but notice that two parallel sections would result from the shadow edges cast on some surface by a relatively narrow object such as a tree trunk. Another possibility concerns motion: successive views of a moving shadow edge. Successive positions of a shadow edge that sweeps across a surface in translatory motion would constitute parallel sections of the surface. Does the visual system take advantage of this fact? Is our ability to analyze parallel surface contours a derivative of an ability to analyze moving shadows? This hypothesis would be supported if we could perceive a surface defined only by a single moving contour that scans across an otherwise invisible surface. In fact, this ability may be demonstrated by a motion sequence of a single contour on a CRT, where each frame presents only a single curve. Note that the moving curve might be interpreted simply as a flexible wire that bends as it translates, or more literally, as a curve in the plane of the screen that changes shape as it moves. But, in fact, there are instances when we interpret the moving contour as a shadow edge sweeping across a 3-D surface (e.g., when the individual curves in figure 1 are presented in succession).

3.3.2 Specular reflections: gloss contours and highlights

Gloss contours, like shadows, are fortuitous, i.e., useful but not necessarily present. They are present only under directional lighting conditions on specular surfaces, when the surface normal lies in the plane defined by the point light source, surface point, and viewer and bisects the angle defined by that configuration. This configuration (the *specularity condition*) is rarely met with planar surfaces but is commonplace for curved surfaces, especially when viewed indoors with multiple lights illuminating the surface. The specularity condition may be met only at an isolated point, causing a *highlight*, or met along a curve, causing a *gloss contour*.

For a doubly curved patch of surface the specularity condition is met at only a point, if at all,

and would only produce a highlight in the image. A gloss contour cannot occur on a surface with nonzero Gaussian curvature in orthographic projection given a point light source.[1] For a gloss contour to occur -- for the specularity to appear not as a point but as a curve -- the specularity condition must be met along a continuous curve on the surface. With orthographic projection and distant light source it is necessary that the contour generator (the locus along which the specularity condition is met) be planar. That plane corresponds to the tangent plane to the surface along the contour generator. Now two results in differential geometry are useful:

> A curve is asymptotic if it lies in a plane everywhere tangent to the surface along the curve.
>
> If the angle between a planar curve and the tangent plane of the surface is constant, then that curve is a line of curvature.

Using the above, we may conclude that the curve across the surface that corresponds to the gloss contour is asymptotic and a line of (least) curvature. Since the asymptotic curve follows a path of zero Gaussian curvature, we have information about the intrinsic geometry in the vicinity. Of importance is the following which hold subject to general position:

> If the gloss contour is curved, the surface is *planar*. This is true in orthographic projection with distant light source. (With nearby objects and perhaps nearby illumination, the surface would not be strictly planar. But in general the surface curvature measured along the contour generator will be small, much less than that measured across the contour generator.)
>
> If the gloss contour is straight, the surface is *cylindrical* when either (a) gloss contours from successive viewpoints are parallel, or (b) if there are multiple light sources (as is common in interior scenes) and multiple gloss contours are parallel.

Thus the specular reflections in the image can tell us not only something of the reflectance

---

1. In real situations we have two ways in which gloss contours may arise. First, extended light sources (such as fluorescent lights, bright windows) will extend point reflections into images of the light sources, which appear as gloss contours if compressed because the two principle curvatures are very different. Secondly, in perspective projection we may have that as the line of sight sweeps across the surface (the projection is not parallel) the angle between the line of sight and the surface stays relatively constant due to curvature of the surface, such as when viewing the inside surface of a cup from nearby. Then if the specularity condition is met at one point in that vicinity, it would be met along a locus. Thus in perspective projection highlights may spread into gloss contours as well.

properties of the surface, that the surface is specular [Beck, 1972], but also something about the surface shape, namely, that the Gaussian curvature is nonzero in the vicinity of a highlight and zero in the vicinity of a gloss contour. The shape of the gloss contour also specifies the intrinsic shape of the developable surface.[1] This does not strictly hold when the surfaces or light sources are near by, and especially when the light comes from an extended, rather than a point, source. Nonetheless, it is instructive to observe the gloss contours on specular surfaces -- they almost invariably follow the least curvature paths on actual surfaces.

3.3.3 Shading contours and terminators

The previous discussion assumes bright, directional light sources. However the specular surface not only reflects the light sources as a highlight or gloss contour, but also acts as a mirror -- the various glossy reflections comprise an image of the surrounds distorted by the geometry of the surface. This is the extreme case of mutual illumination which makes "shape from shading" difficult. The incident illumination is an intractably complex function of the surrounds. But without understanding this illumination, the shape of the surface cannot be solved from the shading.

With the addition of a matte component, the fine details in the reflections are lost, and the gloss contours become less definite. In the limit case of a Lambertian surface there is no specular component and the shading is only a function of the surface orientation relative to the various sources of illumination. For this reason one would expect that the surface orientation would be computed from shading most feasibly, however the illumination is still determined by the surrounds and is still quite unconstrained. Consequently, the computation of shape from shading (where "shape" means local surface orientation) is quite difficult.

Most surfaces are neither totally matte nor glossy so their images present weak highlights and gloss contours -- the distinction between shading and gloss becomes vague. One may postulate, therefore, that shading only constrains the local surface geometry in the manner just described -- the local surface orientation is not computed directly from the shading. Instead, the local surface orientation would be smoothly interpolated between those tangential contours and surface contours along which surface orientation can be solved. The interpolation would be subject to the constraint

---

1. Furthermore, the surface normal coincides with the normal to the plane containing the gloss contour, but to utilize that fact the 3-D curve corresponding to the gloss contour must be determined. That is the topic of section 4.1.

on intrinsic surface geometry provided by the gloss and shading contours. This constraint is naturally described in terms of Gaussian curvature: A highlight indicates positive Gaussian curvature in the vicinity. Similarly, a gloss contour indicates a locus of zero Gaussian curvature.

Constraint on intrinsic geometry is also provided by the shading contours known as *terminators*, surface contours which correspond to paths on the surface along which the light grazes the surface so that points on one side of the contour are illuminated, points on the other side are in shadow. (A terminator is analogous to a tangential contour seen from the light source position.) A strong restriction on the surface shape is provided wherever the terminator is straight in the image: the surface is locally developable (again, assuming general position) and therefore the terminator indicates a locus of zero Gaussian curvature.

4. HOW THE CONSTRAINTS ARE USEFUL

Thus far we have discussed a number of geometrical properties that may be useful in constraining the analysis of shape from surface contours. Instances in which these properties hold in real scenes were described. What remains is to become more specific about why these properties are computationally useful.

4.1 The relation between a surface contour and its contour generator

The current problem is to determine the contour generator Γ in 3-space on the basis of its projection, the surface contour C. To make this problem graphic, observe the 3-D shape of the curve in figure 4. We are asking how that shape may be determined.

The projection will be restricted to be orthographic. This restriction would hold whenever the dimensions of the curve in space are small relative to the distance from the curve to the viewer. Orthographic projection is linear, hence some useful geometrical properties are preserved, notably parallelism.

Now, in determining the shape of contour generators in 3-space we are confronted with a problem wherever the tangent to the contour (its slope) is discontinuous: Is that discontinuity the projection of a discontinuity in tangent along the contour generator, or is the discontinuity due to the adjoining of distinct contour generators on the surface? Since this cannot be answered locally without *a priori* knowledge of the specific surface, we follow the principle of least commitment [Marr, 1976b] and partition the surface contours in an image into their smooth segments.

4.1.1 General position

A number of constraints will be consequences of assuming general position -- that the viewpoint is such that images from nearby viewpoints would not present significant differences in the geometry of the projected contours. By this we rule out viewpoints that cause accidental alignments which mislead. For instance, if a contour C is straight from viewpoint V, then assuming general position, it would be straight from a similar viewpoint -- it is not the case that the contour generator Γ is curved in a plane but that plane is viewed "edge on" so that the image of Γ is foreshortened into a straight line. General position allows one to infer properties of contour generators on the basis of their images, such as smoothness, continuity, and parallelism.

Our first application of general position is as follows. Since the contour C is smooth and

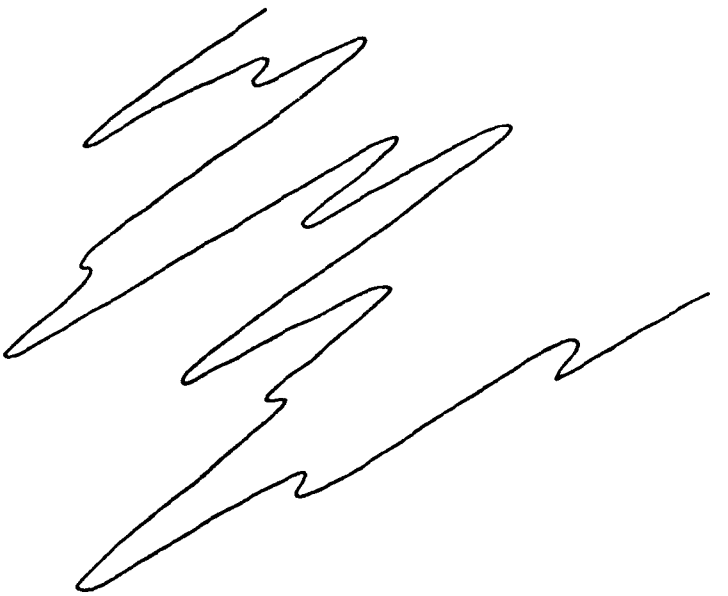


Figure 4. The curve appears to have a specific 3-D shape, as if planar and foreshortened by the slant of the plane relative to the viewer. How is this interpretation derived?

continuous, Γ is smooth and continuous.[1] Furthermore, in general position, nearby and distinct points on Γ project to nearby and distinct point on C. That is, there are no kinks or loops in Γ hidden by the particular viewpoint. In short, assuming general position allows us to consider Γ as a smooth wire in 3-space. Now we consider additional constraints which allow us to determine its shape.

4.1.2 The planarity restriction

If the contour generator Γ is constrained to be planar, the shape of Γ would be completely determined by the equation of the plane containing the curve given its orthographic projection C. Hence the planarity restriction reduces the problem of determining Γ to that of finding the spatial orientation of the plane Π containing Γ.

Since the contour generator Γ is determined once Π is specified, one approach is to impose an *a priori* choice of Π, then examine the shape of Γ that results. That is, one assumes a particular spatial orientation for the plane containing the contour generator. But there do not appear to be any reasonable choices for Π, except for the ground plane, i.e., the horizontal plane defined by gravity. However it is not feasible to assume that all surface contours are projections of horizontal contour generators.

Alternatively, one may make *a priori* assumptions about the shape of Γ in the same spirit as assuming that Γ is planar. Then Π would be a consequence of C and those restrictions on Γ. What restrictions can be reasonably placed on Γ, and how are those restrictions to be phrased? I shall consider two -- symmetry and minimum curvature variation.

4.1.3 Symmetry

Bilateral symmetry is commonly found in nature and usually preserved, at least indirectly, in orthographic projection. We are interested in symmetry, for evidence of symmetry in an image will provide constraint on the shape of Γ. We start with the usual definition of a bilaterally symmetric, planar curve as comprising two loci of points that are reflections of each other across a straight line, the axis of symmetry (figure 5*a*). The symmetric points are equidistant across the axis, the line connecting any two symmetric points is perpendicular to the axis, and all such lines are therefore

---

1. We would like to say something about the smoothness of the surface directly under the contour generator on the basis of the surface contour being smooth, but unfortunately that does not follow from general position as stated. The smooth contour generator may lie along a sharp ridge, for instance.

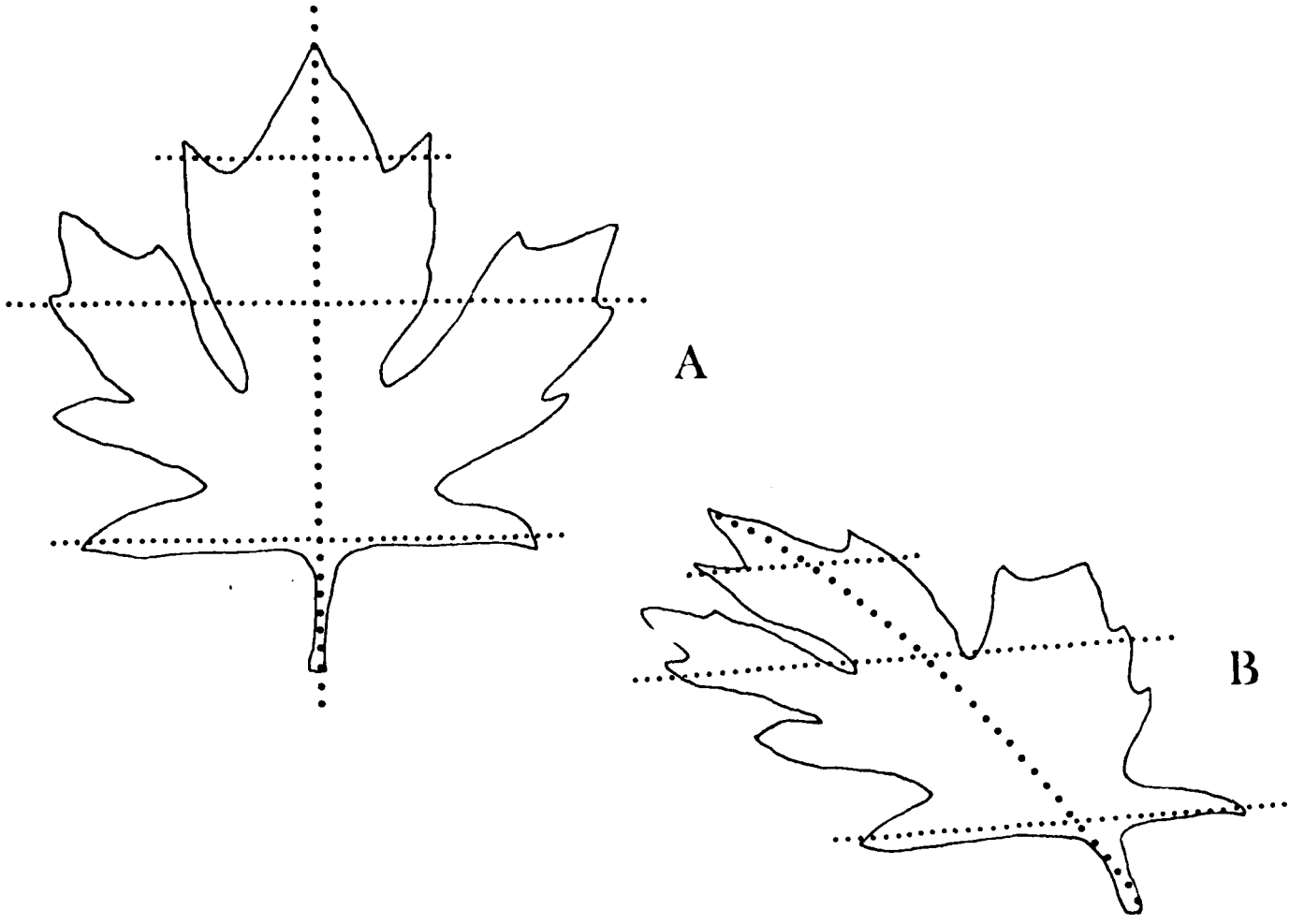


Figure 5. The bilateral symmetry in *A* can be described in terms of correspondence lines which connect symmetric points lying equidistant from a straight line, the axis of symmetry. The parallel correspondence lines are perpendicular to the axis of symmetry. In *B* the correspondence lines connecting qualitatively symmetric segments of the curve are also parallel but make an oblique angle β with the axis of symmetry.

parallel.

In any orthographic projection of this curve, the image of symmetric points are equidistant across the image of the axis, the *correspondence lines* connecting those points are parallel, but the correspondence lines are no longer perpendicular to the image of the axis in general (figure 5*b*). This configuration has been aptly termed "skewed symmetry" by Kanade and Kender [1979]. If a unique line can be found that behaves, in this sense, as the image of an axis of symmetry, then by general position we will assume that the planar curve in space is bilaterally symmetric. That is, we have criteria for detecting bilateral symmetry. When these criteria are satisfied in an image we may assume that it is not coincidental, that it would also be satisfied in an image taken from a different viewpoint -- hence due to actual symmetry. The problem that remains is to detect the images of symmetric pairs of points.

Orthographic projection is linear, hence a number of properties are preserved by the transformation including midpoints, points of inflection, and convexity and concavity [Marr, 1976b]. Marr has shown, in the context of finding the axes of generalized cones, that axial symmetry can be efficiently detected by the *qualitative symmetry* between convex and concave segments, rather than on a point-by-point basis. This extends to the detection of bilateral symmetry, where the correspondence lines between qualitatively symmetric segments would be parallel. The line defined by the midpoints of the correspondence lines would be the image of the axis of symmetry.

Returning to the problem of constraining the shape of the contour generator, the symmetry detected in C constrains Γ to be symmetric and this in turn constrains the orientation of the plane Π containing Γ. Specifically, Π must be oriented relative to the viewer such that, given C, Γ would be symmetric if lying on Π.

This constraint is simply expressed in terms of the *correspondence angle*, the angle in the image between the correspondence line and the projected axis of symmetry (figure 5*b*). Since the correspondence angle is the image of a right angle on the surface, the magnitude of the correspondence angle β constrains the possible spatial orientations for the tangent plane at that point (see figure 6).

In short, Γ is presumed symmetric if an axis of symmetry can be reconstructed from the midpoints of parallel correspondence lines, where the correspondence lines are constructed between qualitatively symmetric segments of C. The correspondence angle then constrains the spatial orientation of the plane containing Γ.

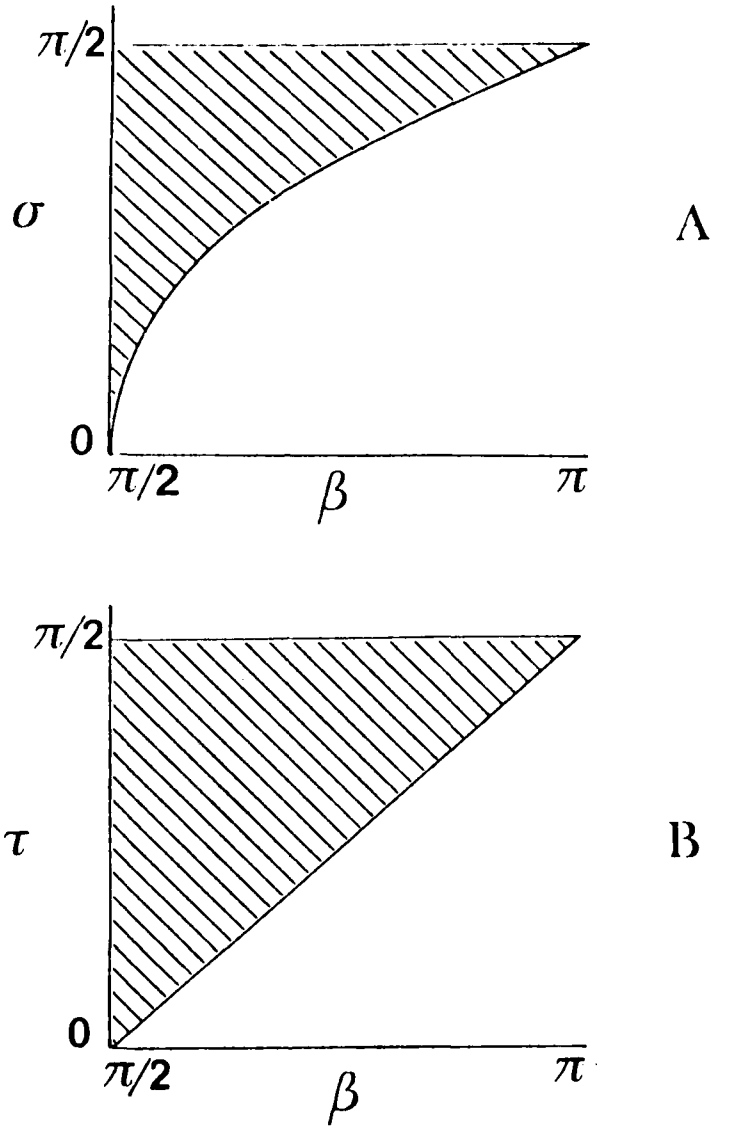


Figure 6. The oblique angle β formed by the projection of a right angle provides some constraint on both the slant σ and tilt τ components of surface orientation relative to the viewer. The possible values of slant and tilt are shown as cross-hatched for correspondence angle β varying from $\pi/2$ to π. Tilt τ is measured relative to one of the contours in the image, and varies from parallel ($\tau = 0$) to perpendicular ($\tau = \pi/2$).

4.1.4 Minimum curvature variation

The curvature of C encodes information about the orientation in space of the contour generator Γ, if Γ is planar and some other restrictions hold. Witkin [1979] has shown that the orientation of the plane Π containing Γ may be estimated on the basis of the curvature along C if we assume that systematic variations in the curvature that resemble foreshortening are due to foreshortening. Then one may choose that plane Π that maximally accounts for the variation in curvature in terms of foreshortening. The following assumptions are sufficient to allow this analysis:

> (a) the possible surface orientations of Π are equally likely,
>
> (b) the tangents to the contour generator are arbitrarily aligned relative to the viewer (they are independent of slant σ and tilt τ), and
>
> (c) the curvature along the contour generator is independent of σ, τ, and the orientation relative to the viewer of the tangent to the contour generator Γ.

The constraint on Γ that results is roughly equivalent to assuming that the variation in curvature along Γ is minimum [Witkin, 1979]. Then the variation in curvature along its projection C may be attributed primarily to foreshortening, whereupon the degree of foreshortening -- hence the orientation of the plane Π containing Γ -- may be estimated. To introduce this, consider the case when Γ is a circle, a planar curve with constant curvature. The orthographic projection C is an ellipse; the curvature along the ellipse varies according to the foreshortening of the corresponding segment of the circle. One may derive from the variance in curvature an estimate of the orientation of the plane containing Γ.

This constraint has been phrased in terms of minimum curvature variation, but Witkin describes it more generally as a problem of signal detection. The "waveform" that we consider is the contour in the image (parameterized in terms of contour curvature). The curvature at any point on the contour consists of two components, one being the curvature of the contour generator at each corresponding point, the other being a "projective component" which increases or decreases the apparent curvature according to the orientation of the given segment of the contour generator relative to the viewer (in the circle example, where the tangent lies parallel to the image plane, the curvature on the ellipse is minimum; where the tangent to the circle is oriented away from the viewer the curvature is greatest). The curvature of the contour generator is treated as noise; the projective component is the signal. Since the projection is orthographic and the contour generator

is planar, the projective component will be regular.

The problem of determining the orientation of the plane containing Γ may be recast as that of estimating the amplitude and phase of a signal of known waveform (the projective component) in the presence of noise (the unknown shape of Γ). The problem can then be solved by seeking to account for as much as possible of the variance in the surface contour in terms of the projective component. The constraint stems from the fact that the processes that determine the shape of contour generators on actual surfaces usually do not impose the same kind of systematic regularity as that imposed by orthographic projection.

4.2 The relationship between a contour generator and the surface

Given the contour generator Γ is a planar 3-D curve, how does the surface Σ lie under Γ? In terms of the wire and ribbon, a primary question concerns whether the ribbon may twist along the wire. More formally, if the plane containing Γ is Π, does the angle between Σ and Π vary along Γ?

A result in differential geometry is that given a curve Γ defined by the intersection of a plane Π and a surface Σ, if the angle between Σ and Π is constant along Γ, Γ is a line of curvature (see, e.g., [O'Neill, 1966, p. 224]). Thus if the contour generator is planar, and that plane intersects the surface with a constant angle, the contour generator is a line of curvature. The next issue is to determine the angle between Π and Σ.

4.2.1 The geodesic and asymptotic restrictions

If the plane Π containing the contour generator Γ is perpendicular to Σ, i.e., Γ is a normal section, then Γ is geodesic. Consequently the surface normal along Γ everywhere coincides with the principal normal to Γ. In essence, the contour generator follows a path on the surface which locally indicates where the greatest curvature occurs. The binormal to the contour generator, being perpendicular to both the principal normal and the tangent, coincides with the direction of least curvature. However all such binormals are parallel, for the tangent and normal along Γ only rotate in the plane Π. Consequently all lines of least curvature are parallel; equivalently, the strip of surface under the contour generator is a cylinder.

The previous discussion considered the case where the contour generator is geodesic; where the angle between Π and Σ is $\pi/2$. If that angle is everywhere zero, then Π coincides with the tangent plane of Σ and the surface normal along Γ coincides with the normal to Π. As mentioned earlier if a curve lies in a plane everywhere tangent to the surface along the curve, that curve is asymptotic,

i.e., a locus of points of zero Gaussian curvature. The importance of the asymptotic restriction is found in gloss contours. The contour generators corresponding to gloss contours in the image correspond to asymptotic curves on the surface. Hence where gloss contours appear we know that the surface is locally developable (likewise, where point specularities occur we also know that the surface must be doubly curved). To some extent we may further understand the surface geometry simply on the basis of the shape of the contour in the image without determining the particular 3-D shape of its contour generator. If the contour is a straight line in the image we cannot tell much, for the surface may be either cylindrical or twisting (like a spiraling piece of paper). But if it is any smooth curve in the image the surface is roughly planar since the contour generator is restricted to be planar and asymptotic.

4.2.2 Parallelism

The discussion thus far has concerned the analysis of surface shape from a single surface contour. This analysis requires that the contour generator Γ may be determined from its image, however the constraint afforded by planarity, general position, symmetry, and constant curvature will not always allow a strong determination of Γ. It is perhaps not coincidental that, in fact, our perception of surface shape from a single, unfamiliar contour is weak when compared to the vivid impression afforded by multiple, parallel contours (figure 1). The basis for the apparently greater constraint from parallel contours will now be discussed; further comments on the psychology of surface contours are given in [Stevens, 1979].

If surface contours are parallel in the image, then by the of general position, their contour generators are parallel. The fundamental issue now concerns the behavior of the surface between the contour generators. In the absence of independent sources of information about the surface such as shading or texture we must make some *a priori* assumption about the nature of the surface between the contour generators. A conservative assumption would be that the surface extends in a "simple manner" between them. This can be formalized by a second form of general position: that the particular positions of the contour generators on the surface are not critical, that if shifted slightly, the contour generators would project qualitatively the same. This is equivalent to assuming that the surface is a cylinder between the contour generators.

We now use the geodesic-asymptotic restrictions from the previous section, and consider two interpretations for the cylindrical surface: Either the surface is (a) curved and the contour generators are parallel geodesics, or (b) flat and the contour generators are asymptotic curves. To

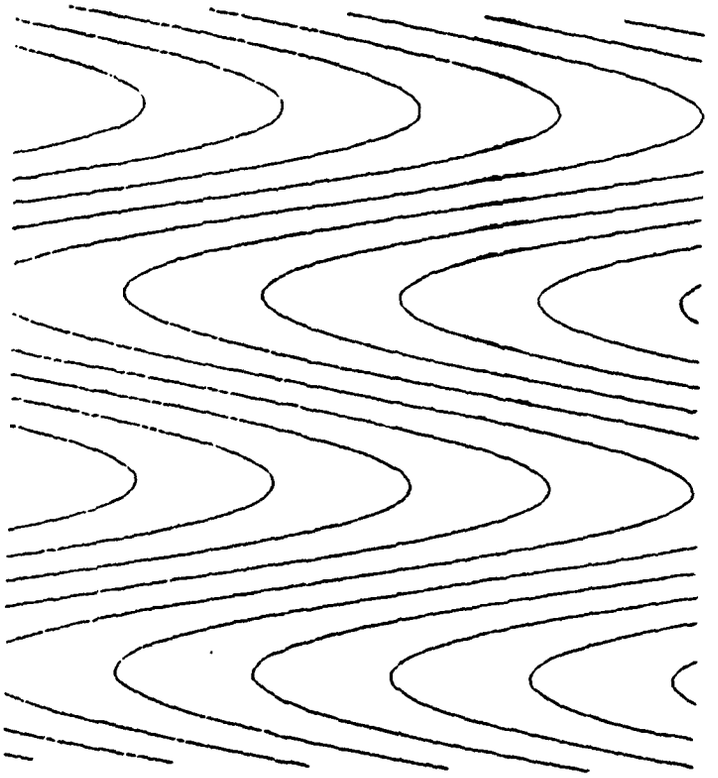


Figure 7. The contours seem to be interpreted as the image of asymptotic curves on a planar surface.

aid in visualizing these two cases, compare figure 1 (geodesic interpretation) and figure 7 (asymptotic interpretation). These effects are discussed in [Stevens, 1979]. Note that in the latter case of asymptotic curves, the parallelism does not provide additional constraint on the surface solution -- the contour generators lie in the same plane. Nor does the shape of each contour generator in the plane; it is as if the curves are merely arrayed on a flat surface. The interpretation of parallel contour generators as geodesics, however, constrains both the local surface orientation and the shape of the contour generators.

4.2.3 Computing parallel correspondence

Recall that the angle between the plane containing the contour generator and the surface is restricted to be constant, hence the contour generator is a line of (greatest) curvature. Also, the lines of least curvature on a cylinder are straight, parallel, and perpendicular to the lines greatest curvature. If a line of least curvature were reconstructed in the image, the angle of intersection that it would make with a surface contour (a line of greatest curvature) would be the projection of a right angle. This angle constrains the local surface orientation, as already demonstrated with regard to bilateral symmetry. In fact, the lines of least curvature can be reconstructed.

In the orthographic image of a cylinder the lines of least curvature would project as straight and parallel, and each would intersect successive surface contours at a constant angle (since the contour generators are parallel). This is illustrated in figure 8. Note that we attempt to reconstruct only the *projections* of the lines of least curvature. This may be achieved by identifying points on adjacent contours whose tangents are parallel and connecting those points by straight lines that are parallel. This may be thought of as bringing points on adjacent contours into *parallel correspondence*. The constructed line representing the image of a line of least curvature will be termed a *correspondence line*. Note that if the surface contours are straight for a portion of their length (figure 9*a*) the tangent to a point P on one contour may be parallel to various tangents on the adjacent contour, however only one choice would result in a correspondence line that is parallel to the other correspondence lines between curved portions of adjacent surface contours (figure 9*b*).[1]

This correspondence is unique in general, and therefore may be used as a constructive criterion

1. Selection of that choice may be accomplished by a local, parallel algorithm similar to that in [Stevens, 1978].

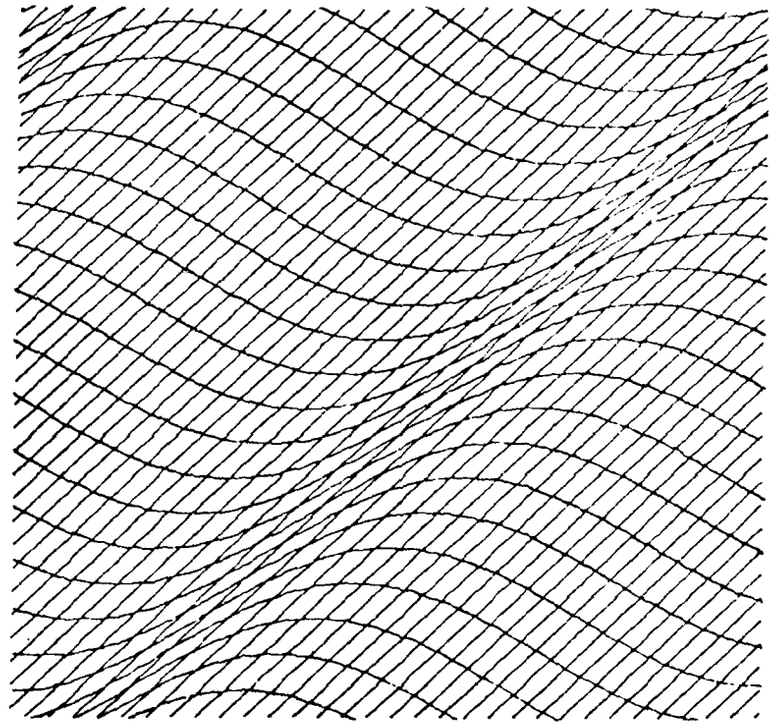


Figure 8. In the orthographic image of a cylindrical surface the lines of least curvature project as straight and parallel, and each intersect successive surface contours at a constant angle. Identifying points on adjacent contours whose tangents are parallel and connecting those points with lines that are parallel establishes *parallel correspondence*, one basis for postulating that the underlying surface is a cylinder (subject to general position).

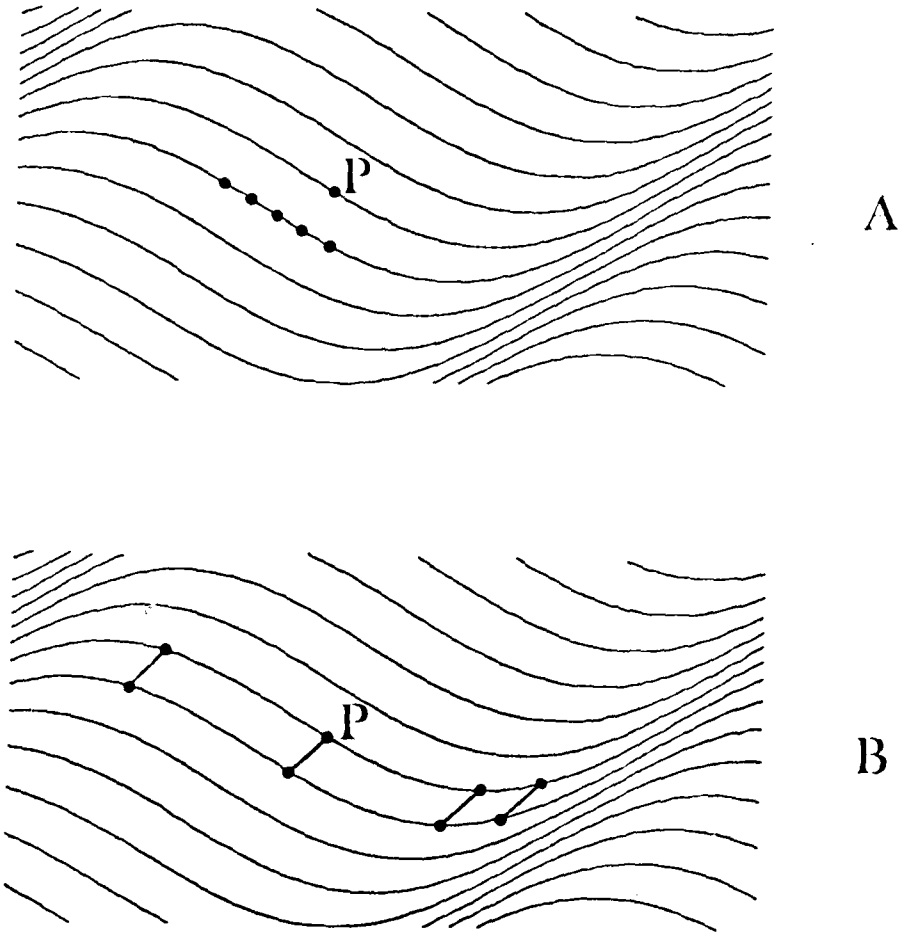


Figure 9. If the surface contours are straight for a portion of their length, as in *a*, the tangent to a point P on one contour may be parallel to various tangents on the adjacent contour, however only one choice would result in a correspondence line that is parallel to the other correspondence lines between curved portions of adjacent contours, as in *b*.

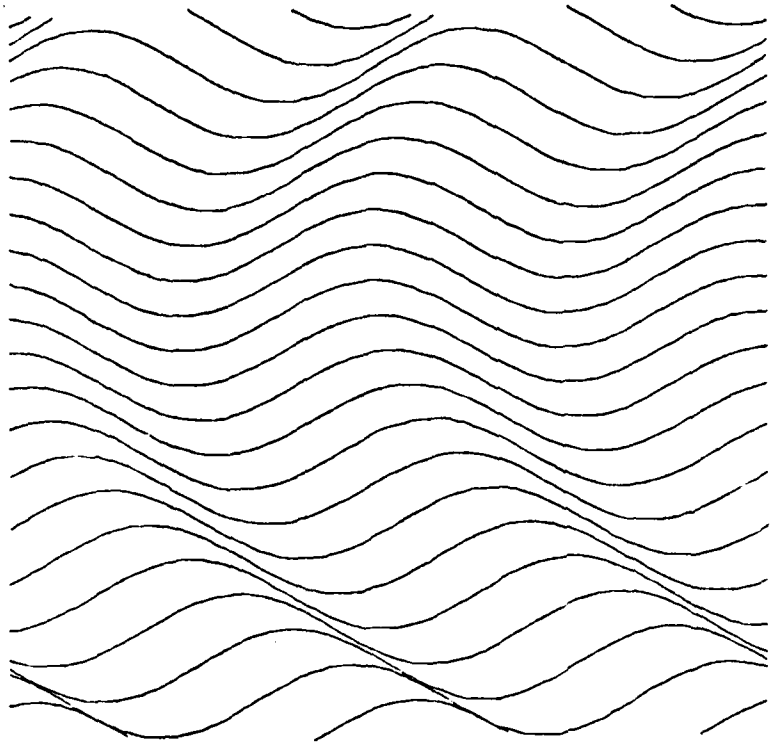


Figure 10. The cylinder restriction is only local, for the parallel correspondence need only be established between adjacent surface contours, and the parallelism between reconstructed lines of least curvature is defined only locally. Consequently the local cylinder restriction may be applied to the surface contours above although the global surface is not cylindrical.

for detecting parallelism between surface contours and for postulating that the surface is a cylinder.[1]

An important consequence of the parallel correspondence is that the surface orientation is necessarily constant along the lines of least curvature (in orthographic projection, as we have been assuming). Thus if the surface orientation were determined along the contour, it can be simply propagated along the correspondence lines to provide a complete, interpolated solution to the surface orientation across the cylindrical surface between parallel surface contours.

We have seen that assuming that the contour generator Γ is planar and that the angle between the plane containing Γ and the surface is constant along Γ restricts the surface under Γ to be a cylinder. Also, for parallel surface contours the two forms of general position together restict the surface to be a cylinder. Consequently, the curvature of the surface is attributed entirely to the curvature of the contour generator, that being a line of greatest curvature.

Note that the cylinder restriction is only local, for the parallel correspondence need only be established between adjacent surface contours, and the parallelism between reconstructed lines of least curvature is defined only locally. Consequently, the cylinder restriction may be applied, for example, to the surface contours in figure 10, where the surface may be approximated locally by patches of cylinders while the global surface is not cylindrical.

4.2.4 Opacity

We now consider the constraint afforded by restricting the surface to be opaque. In general, opacity does not significantly restrict the shape of the underlying surface. However the opacity restriction is important if, as before, the contour generator is assumed to be a line of greatest curvature and the surface under the contour generator is assumed cylindrical. In the following, a geometrical construction will be described that shows how these restrictions constrain the range of orientations to which the parallel lines of least curvature would project. The angle between those lines and the tangent to the surface contour is, again, the projection of a right angle. Thus the opacity restriction is useful in constraining local surface orientation in the same manner as skewed symmetry and parallel correspondence. The restriction imposed on slant and tilt as a function of this angle is shown in figure 6.

---

1. Note that the correspondence is not unique if, for instance, the parallel surface contours are periodic, as in figure 1. One solution in that case is to choose the parallel solution which results in the shortest correspondence lines.

The constraint follows from the fact that if a line of curvature is continuously visible from a given viewpoint, so must an adjacent line of curvature. This can be described geometrically in the following way: The correspondence lines (the projections of lines of least curvature) that connect adjacent surface contours would make no intersections with the surface contours except at their terminations. That is, the situation in figure 11*a* would be disallowed. (Note that in figure 1, where this does not arise, the surface may be transparent nonetheless.) Now, given a single surface contour (the image of a line of greatest curvature on a cylinder) we have some constraint on where an adjacent line of curvature would project, and this in turn constrains the local surface shape.

This constraint is conveniently represented by the Gauss map (see, for example, [Hilbert & Cohn-Vossen, 1952]). A Gauss map is a simple representation of the range of orientations of tangents along a curve. The given curve is mapped to an arc on a unit semi-circle where each point on the curve maps to the point on the semi-circle whose radius is parallel to the tangent to the curve. This is illustrated in figure 11*b*. Observe how tangents at various points P map to corresponding points on the semi-circle.

The next step is to use the Gauss map to represent the range of possible orientations of the correspondence lines. Let that orientation be α, which maps to a single point P on the semi-circle (that point P whose radius has the orientation α). In figure 12 three choices for α are shown which are consistent with the surface being opaque. Now, the constraint that the correspondence lines not intersect the surface contours equates to the restriction that the point P (the mapping of orientation α) not lie on the arc of the semi-circle already covered by the surface contour. The degree of constraint imposed by the opacity restriction depends on the surface contour. In figure 13*a* the shallow contour maps to only a short arc, and the correspondence lines could have a large range of orientations. But in figure 13*b* the correspondence lines are restricted to a narrow range of orientations.

Given that the correspondence lines are the projections of lines of least curvature which on a cylinder are identically the binormals to the plane containing the lines of greatest curvature, the orientation to which the correspondence lines projects provides us with the tilt component of surface orientation for the plane containing the given curve.

4.3 Criteria governing the tangential/surface contour decision

Earlier we discussed the distinction between tangential contours (silhouette boundaries along which the line of sight grazes the surface) and surface contours, noting that surface contours include

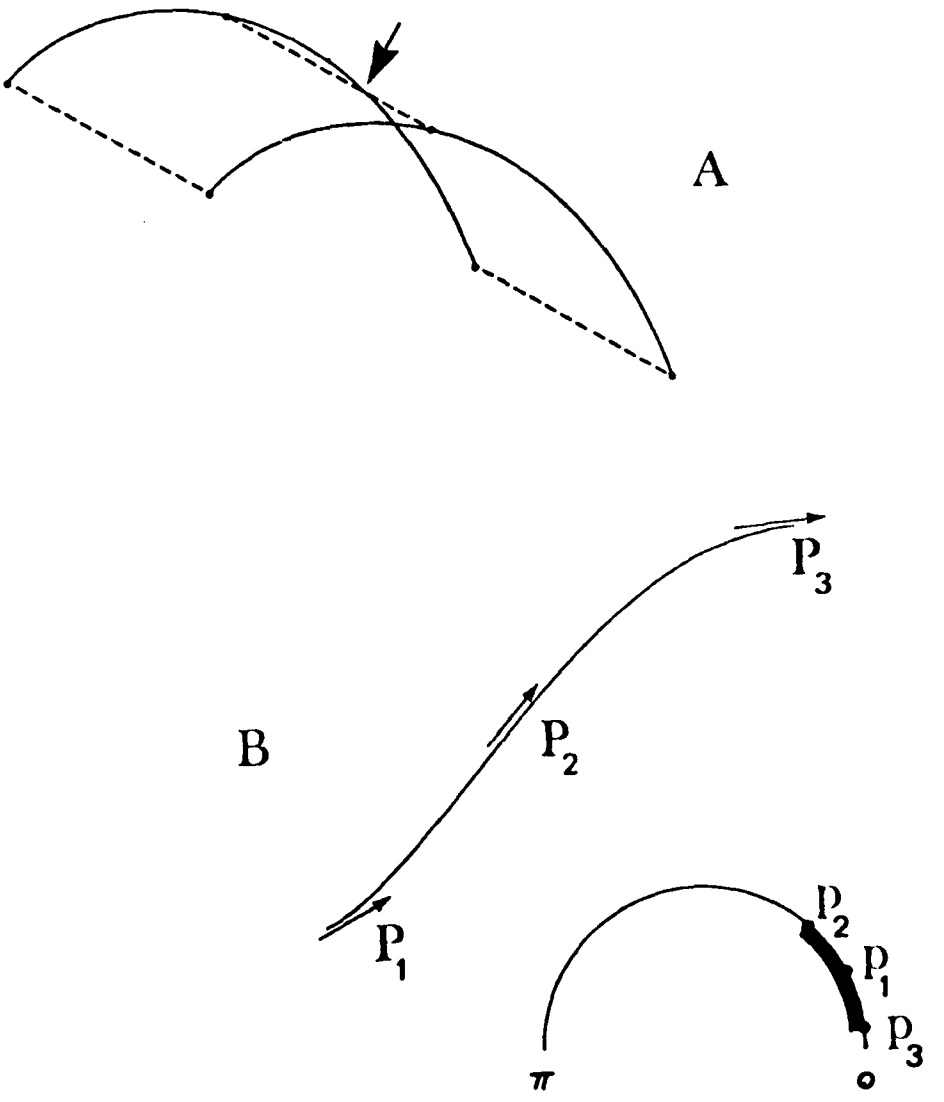


Figure 11. The opacity restriction disallows the correspondence lines (the projections of lines of least curvature) that connect adjacent surface contours to intersect the surface contours except at their terminations. That is, the situation in *a* is disallowed. Opacity provides some constraint on the relation between a contour generator and the underlying surface. Towards representing this constraint, we represent the surface contour by its Gauss map onto a semi-circle, as in *b*.

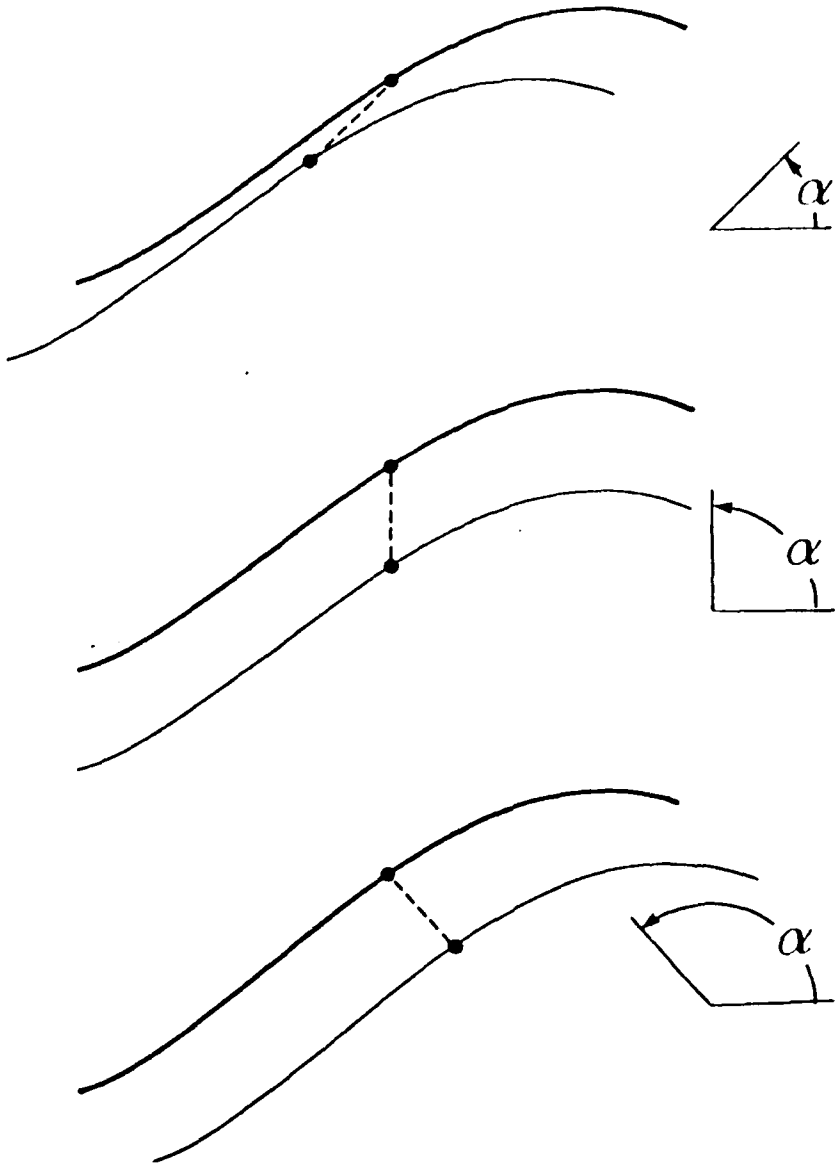


Figure 12. The surface underlying the contour (heavy line) is assumed to be a cylinder, and the problem is to determine the orientation α to which the lines of least curvature would project. Three examples of α are shown above. The opacity restriction places some constraint on α.

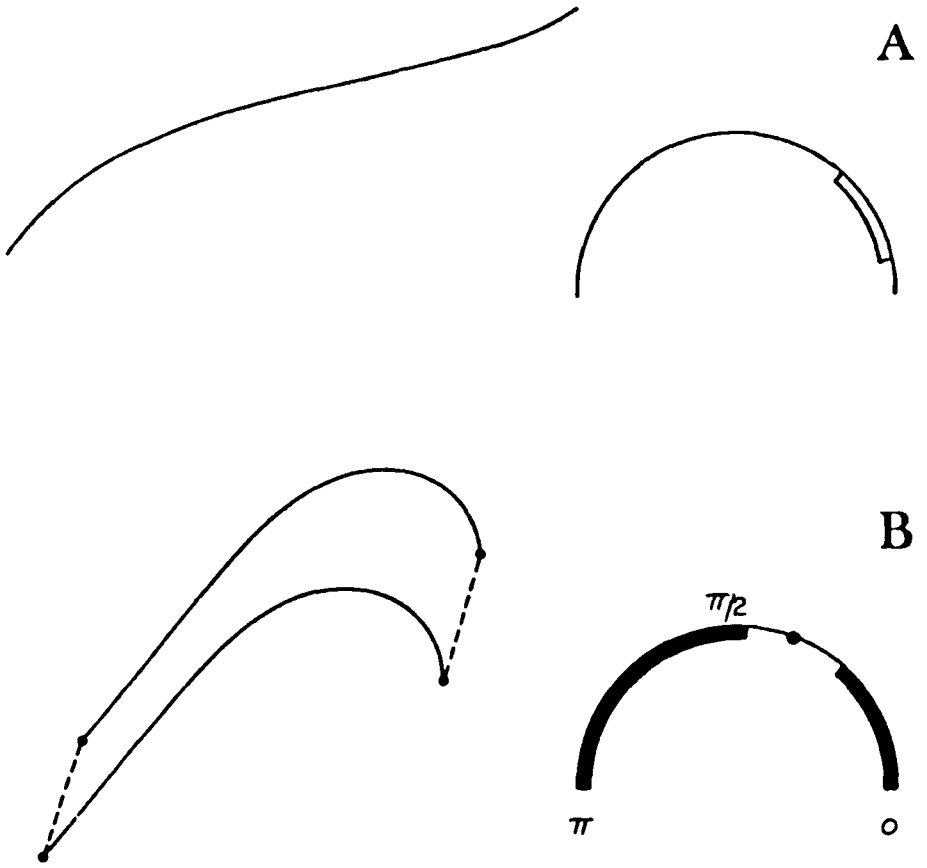


Figure 13. The image of the lines of least curvature map to a single point on the Gauss map. If opaque, that point cannot already be occupied by the mapping of the surface contour. In *a* the surface contour is a shallow curve which maps to a small arc on the Gauss map. This does not strongly constrain the possible orientations of the correspondence lines (the projected lines of least curvature). But in *b* the curve covers much of the Gauss map, hence the orientation of the lines of least curvature is strongly constrained. One choice of that orientation is shown, and the position of an adjacent, parallel surface contour is drawn. The opacity restriction then provides constraint on surface orientation by the oblique correspondence angle.

silhouette boundaries that are not tangential contours. Marr [1976b] has delineated properties of the silhouettes of generalized cones (whose boundaries are tangential contours) -- surfaces whose shape can be recovered from their silhouettes. The silhouette of a generalized cone exhibits qualitative symmetry; where the correspondence lines connecting symmetric segments of the contour would be perpendicular to the axis of symmetry. For instance, the symmetric silhouette in figure 2*a* is generally interpreted as a vase-like object, and the contours are seen as tangential contours.

Similarly, geometrical criteria can be given which indicate that a contour is a surface contour. (Note that non-geometrical means also exist, e.g., determining that the corresponding contour generator is a shadow edge, or a gloss contour or a discontinuity in surface texture) Two geometrical criteria are suggested by the preceding discussion. First consider qualitative symmetry where the correspondence lines are not perpendicular to the axis of symmetry (as just discussed in the case of bilateral symmetry) but oblique to the axis (as in figure 5*b*). When achieved, this skewed symmetry suggests a surface contour, as opposed to a tangential contour, interpretation. Secondly, if parallel correspondence between contours can be achieved, as in figure 3*b*, those contours can be interpreted as surface contours.

5. SUMMARY

1. The analysis of the shape of a surface from surface contours may be decomposed into two problems: reconstructing the corresponding 3-D curves (the *contour generators*) and determining their relation to the surface. This decomposition separates the problem of determining the projective geometry from that of determining the intrinsic geometry.

2. The first problem is constrained by general position, planarity, symmetry, and minimum curvature variation.

3. The second problem is reduced by assuming the angle between the surface and the plane containing the contour generator is constant. Then if that angle is a right angle, the contour generator is geodesic; if the angle is zero, the contour generator is asymptotic. In either case the contour generator is also a line of curvature. Since it is also planar, the surface is locally a cylinder.

4. We also arrived at the cylinder restriction in the case of parallel surface contours, given the two forms of the principle of general position. The opacity restriction is also useful, given the planarity and geodesic restrictions, in understanding how the surface lies under a contour generator.

5. We have considered instances when the various constraints are valid. Surface markings on synthetic and biological objects and the edges of cast shadows are often geodesic and planar. Gloss contours are asymptotic and planar, at least in the case of distant light sources and orthographic projection. Hence if the contour generator can be reconstructed as a curve in 3-D, the surface orientation along the curve can be computed subject to either the geodesic or asymptotic interpretations.

6. Constraints on the intrinsic geometry are also provided by surface contours even if the contour generator is not well determined in space: Gloss contours, highlights, and shading edges tell us of the local Gaussian curvature in some cases.

PART II: TRANSPARENCY

Whitman Richards and Andrew P. Witkin

1. INTRODUCTION

1.1 The problem

The intensity of light reaching a point in an image such as the retina depends upon many parameters of the scene. These include the strength of illumination, reflective properties of the viewed surface, and the orientation of the surface with respect to the viewer and the illuminant. To interpret an image, these confounded factors must be sorted out. Important steps in this direction have been taken by Land and McCann [1971] and by Horn [1975]. Land and McCann have developed a method for differentiating the contributions of illumination and reflectance for simple scenes, while Horn has shown how to isolate the contribution of surface orientation.

For natural scenes, still another factor may confound the interpretation of image intensities: when a partially transparent object occludes the scene behind, then the resultant image is a combination of the properties of the transparent object as well as its background. How can the object and its background be isolated from the available image intensity distributions? Such situations occur frequently, not only when scenes are viewed through haze, rain, or dirty windows, but also when a tree or bush is interposed between the object of interest and the eye. In such situations, the image content on the retina then becomes a projection of both the transparent occluding medium and the scene behind. How can these two components of the image be separated so that the true scene content can be deduced?

1.2 Extending the definition of transparency

When viewed at a distance, a dust cloud appears to be a homogeneous partially transparent medium. But if we get close enough to resolve the individual dust particles, we find that it is composed of opaque elements suspended in a completely transparent medium. Does this mean that the cloud is no longer partially transparent beyond the point at which its structure can be resolved? Clearly any collections of opaque suspended elements -- i.e., "screens" -- become transparent media whenever the elements cannot be resolved. The distinction between a "screen" and the classical

transparent "film" is merely one of the scale of resolution. The general form of transparency, therefore, includes the separation of such "screens", e.g. trees or rain, from their backgrounds. The homogeneous film is merely a special limiting case.

A general definition of transparency is therefore as follows:

> **A surface (or medium) will be said to be transparent if the projected image content depends both upon the medium as well as its partially occluded background.**

This definition allows for the spatial (or temporal) frequency content of the image to be altered by interposing a transparent object or medium, as well as alterations by intensity, color, or contrast changes. Quite simply, a transparent medium is a filter that passes some of the more distant scene content, as well as contributing to the scene. The general problem is to assess the properties of this filter and its background from image information alone.

1.3 Historical: When can transparency be deduced from image information?

Early experimenters readily recognized that perceptual transparency does not correspond to physical transparency [Metelli, 1970]. For example, air is physically transparent, but we do not see air as a transparent medium unless it is a misty day. A clear pane of glass is another example of a physically transparent medium that will not be "seen" as transparent unless the pane has a dirty film or a reflection on it. Thus, physical transparency is not a necessary condition for perceptual transparency. Nor is it sufficient, for consider the years of successful experimentation on transparency using opaque cardboard cut-outs or printed illustrations as in Figure 1 [Metelli, 1974]. Does this mean therefore that physical transparency cannot explain perceptual transparency?

Although air (or aqueous humor) is physically transparent, and not seen, should this observation exclude us from considering the perception of transparency as an attempt to determine whether a filter has been interposed between retina and a portion of the external scene? Obviously a clear pane of glass will not appear transparent because no information about the presence of the glass is available in the retinal image. Clearly, if transparent media are to be "seen" by a visual system, then information about transparency must be available in the retinal image. Understanding the perception of transparency must therefore begin by understanding what information is potentially available in the image. Such an analysis is a first step toward developing a

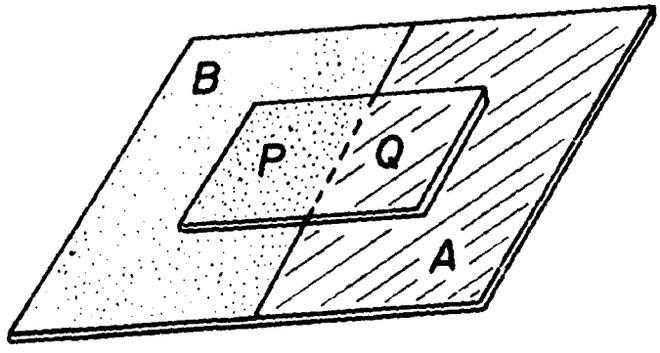


Figure 1. A typical transparency arrangement.

computational theory of transparency (see Marr & Poggio, 1977).

2. BEGINNING A COMPUTATIONAL THEORY

2.1 Measures of "transparency"

Although the basic problem is to deduce and describe the nature of a transparent medium from image information alone, a first step is to determine how we are to characterize the properties of the transparent medium. Relevant descriptions include the spatial frequency content, the color, the reflectance, and the transmittance of the medium. Our analysis will be restricted to untangling the reflectance and transmittance of the transparent medium from the reflectance of its occluded background. Spectral information will also be considered, but in a subsequent section.

2.2 Relations between image intensities, reflectances and transmittance

If no texture or temporal information is available in a region of an image, then transparency cannot be deduced solely from image intensities lying only within the transparent region. Under these conditions, the region (such as P in Figure 1) will be totally homogeneous. Under uniform illumination, we can designate the image intensity in the region (P) as $I_{avg}=I_P$. This average intensity image will be determined by the reflectivity ρ_s of the transparent medium (P) and its transmittance, τ. Because some light from the background B also contributes to the observed image intensity in region P, we must consider its intensity I_B, which will be the product of the reflectance of the background ρ_B and the illumination E common to all surfaces in the scene. The combined intensities I_{avg} seen be the viewer will then be

$$I_{avg} = E \cdot \rho_S (1-D) + E \cdot D \cdot \tau \cdot \rho_B \qquad (1)$$

where D is the areal density of the transparent regions of the medium and (1-D) is the density of the reflecting particles of the medium. Thus, if the medium is considered as a screen, the fraction of the screen that reflects light is (1-D), whereas D is the fraction of the screen that is occupied by "holes" and is completely transparent. Of course, as the scale of the screen becomes finer and finer, eventually both the reflecting particles and the holes between them become invisible to the observer, who will see only a homogeneous medium.

Returning now to equation (1), the problem is to determine the reflectance ρ_S and transmittance τ of the medium P (which may be a fine screen). However, in region P, the only information available in the projected (retinal) image is I_{avg}. Clearly we cannot recover the reflectance ρ_S and transmittance τ from a single equation of one known (I_{avg}) and the five unknowns ρ_S, τ, E, D and ρ_B.

The next obvious option is to consider two adjacent regions, such as P and B in figure 1. If it were known that B is indeed the "background" alone, while P is the mixture of background and an interposed screen (or some other medium), then the image intensity $I_B = E \cdot \rho_B$ is known for this region B and we now have:

$$I_{avg} = E \cdot \rho_s (1-D) + D \cdot \tau \cdot I_B \tag{2}$$

where the first term on the R.H.S. is the light reflected off the surface of P and the second term is the light from B transmitted through P. Since I_{avg} and I_B are now known, only D, ρ_s, E and τ remain as unknowns. However, it is still impossible to solve for ρ_S and τ and thus the properties of P still cannot be separated from its background B. Thus, two homogeneous regions, one of which is known to be the background still do not provide enough information to determine whether the adjacent region (P) is transparent or opaque.

If, however, the areal density of the "holes" in P can be estimated independently (such as by a higher resolution process), then presumably the image intensity $I_s = \rho_s \cdot E$ of the opaque portion of the screen can also be measured by the same process. Thus,

$$I_{avg} = I_s (1-D) + D \cdot \tau \cdot I_B \tag{3}$$

and since D, I_s, I_{avg}, and I_B are available from the image, the transmittance of the region P can be deduced (but not its reflectance!):

$$\tau = \frac{I_{avg} - I_s (1-D)}{I_B \cdot D} \tag{4}$$

In many cases, higher spatial resolution may not be available to provide independent measures of the hole density D and the intensity I_s of the opaque portion of the screen. (This is the limiting condition where the "screen" is seen as a transparent film or filter.) Metelli [1974] has noted that for this condition, two independent backgrounds are needed to assess the transmittance of the transparency. As in figure 1, let A be the second background whose image intensity is measured as

I_A and let Q be the region of the transparent medium lying over A, with image intensity I_Q. Then if P+Q is indeed a physically homogeneous transparent region, both D, ρ_s and τ will be constant over P+Q. We thus have two equations with four unknowns, E, ρ_s, D and τ:

$$I_P = E \cdot \rho_s (1-D) + D \cdot \tau \cdot I_B$$

$$I_Q = E \cdot \rho_s (1-D) + D \cdot \tau \cdot I_A \qquad (5)$$

However, by treating the product $D \cdot \tau$ as one unknown, we see that the remaining unknowns cancel to yield:

$$D \cdot \tau = \bar{\tau} = \frac{I_P - I_Q}{I_B - I_A} \qquad (6A)$$

where $D \cdot \tau = \bar{\tau}$ is the average apparent transmittance of the medium P+Q. Note that as before, extended illumination has been assumed in the derivation of the relation (6A). If the lower surface B+A is diffuse and is illuminated through the transparency P+Q, then the true average transmittance is the square-root of the R.H.S. of (6A), since in that case the light reaching the viewer has passed through the transparent medium twice. Considering this uncertainty about the illuminant, the true value of the transmittance τ cannot be determined. Recognizing however that τ lies between 0 and 1, we note that a weaker condition for P+Q being transparent is

$$1 \geq \frac{I_P - I_Q}{I_B - I_A} > 0 \qquad (6B)$$

Metelli [1974] proposes that whenever equation 6B is satisfied, then P+Q will appear transparent.

In summary, if high spatial resolution is not available, then the calculation of the average transmittance of a "region" in the image requires the inspection of four mutually adjacent regions in the image. This is the minimum configuration, which reduces to a local "x" intersection for homogeneous fields. In this case, the average (relative) reflectance ρ_s of the screen can also be deduced and is found to be

$$\rho_s = \rho_s \cdot D = \frac{I_B \cdot I_Q - I_P \cdot I_A}{(I_B - I_A) \cdot E} \qquad (7)$$

Note that only relative reflectance of the transmitting surface can be found because the illuminant strength E is still undetermined.

2.3 Apparent transmittance, τ^* and contrast

Equations (4) and (6) gave two different values for transmittance, depending upon whether the scale of the image measurement is smaller (equation 4) or much larger (equation 6) than the hole size of the screen. Clearly where the image scale is finer than the hole size, then the transmittance will equal that of the hole, τ, whereas when the image grain is coarse, then the observed transmittance will be the average τ. What, then, is the appropriate measure for the intermediate case?

As a screen becomes finer and finer, so its holes cannot be resolved (as if the observer had moved away), then the transmittance goes from τ to $\bar{\tau}$, with the latter corresponding to the case where the screen becomes a homogeneous transparent film. Clearly, the criterion for deciding whether the region P+Q should be considered a screen or a film depends upon whether the "holes" are visible or not. Such visibility depends upon the convolution of the image operator with the image values, but can be simply regarded as the effective contrast of the step across the edge of an opaque element to the "hole" in the image. Let ΔH be the true intensity change across this border, and ΔH^* be the image value (i.e., the result of convolution with the transfer function of the imaging process). Then the apparent transmittance τ^* can be defined as

$$\tau^* = \bar{\tau}\left(1-\frac{\Delta H^*}{\Delta H}\right) + \frac{\Delta H^*}{\Delta H}\tau \tag{8}$$

Equation (8) can be rearranged to the form:

$$\tau^* = \bar{\tau} + \frac{\Delta H^*}{\Delta H}(1-D)\bar{\tau} \tag{9}$$

Because of resolution losses in the imaging system, $\Delta H^* \leq \Delta H$. Therefore as the contrast between the elements and "holes" in the screen goes to zero either by resolution losses or by $\Delta H \to 0$, the apparent transmittance τ^* will approach the average transmittance $\bar{\tau}$ described by equation (6).

2.4 Imposing natural constraints: joints and "cracks"

Although at least four contiguous regions as in Figure 1 are required for deducing transparency from homogeneous fields using image intensity alone, the configuration is not yet minimal. If the two backgrounds A and B are different structures (as opposed to region A being merely a painted portion of B), then there may often be a crack or joint between A and B where they abut. Of course, this "crack" will continue into the region P+Q, creating the boundary between the

transparent overlay. The crack itself, therefore, is a third intensity profile available in the image, and its image intensities can be used instead of region A (or B).

The computational advantage of such "cracks" when they appear is that, usually, almost no light escapes them. This means that a "crack" seen through a transparent medium gives approximately the pure reflective component of the medium, with negligible transmitted contribution from the background. That is, intensity profiles of most "cracks" and joints approach a contrast step of 100% for widths that are small in relation to the depth of the crack into the surface. Such cracks can generally be identified in the image by their intensity profiles. Once isolated, then the contrast step in the background ΔC can be taken reliably as near 1.0. This useful natural constraint upon images may be exploited if the transmittance and reflectance relations for the transparent medium (equations 6 and 7) are recast in terms of contrast.

Let ΔC^* be the apparent contrast across the AB border:

$$\Delta C^*_{AB} = \frac{I_B - I_A}{I_A + I_B} \tag{10A}$$

and ΔC^*_{PQ} be the corresponding contrast for PQ:

$$\Delta C^*_{PQ} = \frac{I_P - I_Q}{I_P + I_Q} \tag{10B}$$

Substituting (10A and B) into equation (6) we find the average "film" transmittance to be

$$\bar{\tau} = \frac{\Delta C^*_{PQ}}{\Delta C^*_{AB}} \cdot \frac{I_P + I_Q}{I_B + I_A}$$

or

$$\bar{\tau} = \frac{\Delta C^*_{PQ}}{\Delta C^*_{AB}} \frac{(1 + \frac{I_Q}{I_P})\, I_P}{(1 + \frac{I_A}{I_B}) I_B} \tag{11A}$$

Recasting equations 10 in terms of I_Q/I_P and I_A/I_B and substituting in the above:

$$\bar{\tau} = \frac{\Delta C^*_{PQ}}{\Delta C^*_{AB}} \frac{(1 + \Delta C^*_{AB})}{(1 + \Delta C^*_{PQ})} \frac{I_P}{I_B} \tag{11B}$$

For highly transmitting media, such as "thin" films, haze, smoke or rain, equation (11A) is useful because as $I_P \to I_B$, the transmittance is determined by the contrast reduction alone. In particular, for a "crack", $\Delta C^*_{AB} = 1$ and

$$\bar{\tau} = 2 \cdot \Delta C^{*}_{PQ} / (1 + \Delta C^{*}_{PQ}) \tag{12}$$

or to a crude first approximation (since ΔC^{*}_{PQ} also approaches 1) the transmittance equals the maximum contrast in the region P+Q. Thus, for a textured background (that include "cracks" where $\Delta C^{*}_{AB}=1$) and a homogeneous transparency, the measured distribution of contrasts can be used to estimate transmittance. Clearly because transparency of necessity must reduce the contrast in the scene, the distribution of gray levels must become narrower. Equation (12) is just a special instance of this result.

Equations (10) can also be combined with equations (6) to determine the relation between the average reflectance $\bar{\rho}$ of the transparent medium and apparent contrast in region P+Q:

$$\bar{\rho}_{PG} = (1 - \Delta C_{PQ}) \cdot I_P / E \tag{13}$$

Note again, that without some estimate of the illuminant strength E, the absolute reflectance of P+Q cannot be calculated from image intensities. However, if evidence is available that a region B has maximum reflectance of 1, then since $I_B = \rho_B \cdot E$, with $\rho_B = 1$ the reflectance ρ_{PQ} of P+Q can be estimated. Two properties of images can aid in this estimate: 1) the spectral reflectance of B and the maximum intensity adjacent to the perimeter of P+Q *provided that this intensity falls upon the expected distribution for gray levels.*

3. TRANSLUCENCY

In a still more general case of achromatic transparency. the screen or film will have internal scatter, creating a translucent surface if the holes of the screen are not visible in the image. Clearly the scattered light will reduce contrast. If the translucent medium lies on top of a background of reflectance ρ_B, then the apparent reflectance of the translucent surface will be

$$\rho' = \rho_0 + \frac{\tau^2 \rho_B}{1 - \rho_0 \cdot \rho_B} \tag{14}$$

where τ is the internal scatter of the translucent medium and ρ_0 is the reflectance of the medium with a completely absorbing background [Kubelka-Munk, 1931, 1954]. The image intensity I_P will now be

$$I_P = E\bar{\rho}_0 + E\tau^2 \rho_B / (1 - \rho_0 \rho_B)$$

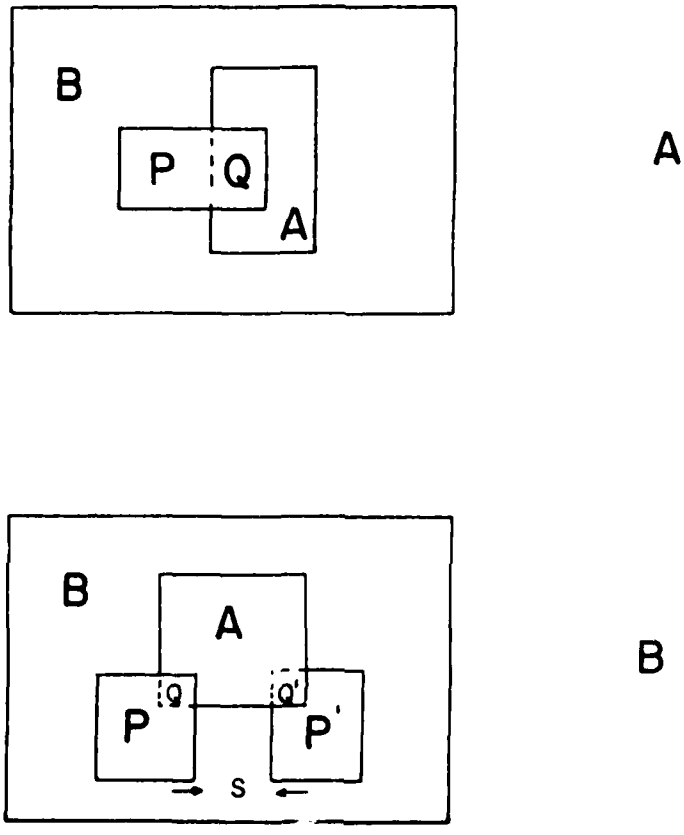


Figure 2. Transparency arrangements used in the present experiments. The upper configuration was used to test Metelli's proposal. The lower configuration illustrates the local nature of transparency judgements. As long as distance s exceeds about one-half degree, the panels P + Q and P^1 + Q^1 can be made to appear transparent although the intensity relation violate the rules for physically transparent objects.

A similar equation for regions Q and A yield the following solution for τ^2:

$$\tau^2 = \frac{(I_P - I_Q)(1-\rho_o\rho_B)(1-\rho_o\rho_A)}{(I_B - I_A)\left((1-\rho_o\frac{(I_B\rho_A - I_A\rho_B)}{(I_B - I_A)}\right)} \quad (15)$$

Although I_P, I_B and I_A are available from the image, ρ_o, ρ_B and ρ_A are not. Hence τ^2 cannot be determined. (However, note that for all ρ's, the quotient always satisfies the relation $0 \leq \tau^2 < 1$ for a transmitting medium.) In the special case where $\rho_o \to 0$, then $\tau^2 = (I_P - I_Q)/(I_B - I_A)$ as before.

Now consider the possibility that a "crack" or edge can be identified in the image (either by occlusion or by the distribution of gray scale values in region P). Then $\rho_A = 0$ and $I_A = 0$, simplifying (15) to

$$\tau^2 = \frac{I_P - I_Q}{I_B}(1-\rho_o\rho_B) \quad (16A)$$

This can be recast in terms of the contrast ΔC_{PQ}. For small ΔC_{PQ} values, P=Q and

$$\tau^2 = \Delta C_{PQ}\cdot\frac{2\cdot I_P}{I_B}(1-\rho_o\rho_B) \quad (16B)$$

An especially interesting case of (16B) is when $\rho_B \to 1$ and where the translucent medium is "white". Then the absorption is trivial and $\tau = (1-\rho)$ and $I_P = I_B$. Substitution in equation (16B) yields

$$\tau \approx 2\Delta C_{PQ} \quad (16C)$$

Thus, the apparent translucence yields twice the image contrast. This is the case when a piece of white onion-skin paper lies over a black line on a surface (or if one views a printed sheet from the back side). Note that equation (16C) is the same as equation (12) for low contrasts between P and Q. Films of low transmittance can therefore appear equivalent to translucent surfaces, with the transmittance approximated by twice the maximum image contrast provided that I_P roughly equals I_B.

To determine an upper bound on the validity of equation (16C), consider the case of a translucent surface with its backing removed. Then ρ_B may approach 1 for a "white" sheet and extended illumination that includes back illumination of the translucency. Letting the reflectance of the medium be ρ_o as before, $I_P = \rho_o \cdot E$ and $I_B = 1 \cdot E$, and equation (16C) becomes

$$\tau = 2\cdot\Delta C_{PQ}\cdot\rho_o(1-\rho_o). \quad (16D)$$

(The exponent for τ is omitted because light passes through only once.) Because $\Delta C \cdot \rho_0 \leq 1$ and because $0 \leq \rho_0(1-\rho_0) \leq .25$, the maximum possible value of τ is 0.5 if ρ_B is a "white" medium. Equation (16C) thus cannot be valid if it yields a transmittance greater than 0.5. A conservative position for its useful upper bound will be to make inferences from equation (16C) only if the apparent transmittance comes to less than 0.25.

4. ACHROMATIC SUMMARY

Transparency and translucency impose different relations between transmittance and image intensities, even if the illumination conditions are the same and known. However, for both transparency and translucency in the limiting case where maximum contrast can be estimated in the region overlaid by the transparency (or translucency), the average transmittance of a homogeneous film will approximate the maximum contrast for highly transmitting media. Haze, smoke, rain or mist meet these conditions. A reasonable image measure for transmittance is thus the apparent contrast reduction. This is our counterproposal to Metelli's theory of transparency. The contrast reduction measure offers a simple approximation to the exact physical solution. It holds exactly for highly transmitting media and offers close approximations for such natural transparencies as haze and rain, which approach those limiting conditions. For media with low transmittances (below 25%), twice the maximum contrast reduction will provide a crude estimate of translucency or transmittance.

Unfortunately, the absolute reflectance of the transparent (or translucent) medium cannot be easily estimated from image information alone, without some knowledge of the illuminant strength. If, however, the reflectance of only one portion of the adjacent background can be estimated [Richards, 1979], then the reflectance of the transparent overlay can be determined.

5. SPECTRAL CONSIDERATIONS

Under achromatic conditions, a most reasonable estimate of transmittance, based upon image intensities alone, is to determine the maximum contrast and set this equal to the transmittance. This determination can be reinforced by evidence obtained from "scratches" or cracks" that cross the occluding border. Such a solution is based upon the assumption that both the translucent or transparent media as well as their backgrounds have flat spectral distributions over wavelengths for

both reflectance and transmittance. In the natural world, however, this is an uncommon occurrence. How then are the transmittance and reflectance of a screening surface affected when spectral factors are considered?

5.1 Why predictions based upon additive mixtures fail

If P+Q is a screen, and if the illuminant is extended, with a flat spectral distribution, then the chromatic content of (P+Q) will be merely the mixture sum of the light reflected off (P+Q) plus the light transmitted through the "holes" coming from the surfaces B and A below (τ is now assumed to have a flat spectral transmittance over wavelength). If the colored image is assessed by a trichromatic system similar to man's, then the color content of P and Q can easily be deduced from the fact that all mixtures of two lights might be on straight lines in a chromaticity diagram [Wysecki and Stiles, 1967].

In figure 3a, let the x,y (chromaticity) coordinates of the reflecting surface (P+Q) be at point ρ, locations B and A respectively corresponding to the two backgrounds (if A=0, then its location is W.) Then all mixtures of transmitted and reflected light will be on the two loci ρB and ρA, at least for the case where the "holes" in the screen have an flat transmittance function over wavelength. Thus P and Q must lie on those loci, with their distance from ρ proportional to D/(1-D) times their tristimulus sums where D is the hole density.

The outer contour in the CIE diagram denotes the spectral limits of all physically realizable x,y values. Thus ρ, B, and A must all lie within the enclosed boundary. This constraint does not preclude the possibility, however, that P and Q can both occupy positions on the opposite side of white (W) from B and A. For example, P and Q might be greenish blue and bluish green, while B and A would be yellowish green and orange. Since such stimuli are seen as four opaque regions, rather than as a transparent situation, we know that the human observer does not consider homogeneous transparent surfaces as screens that obey additive color mixture laws.

Perhaps one obvious explanation for the human's failure to treat transparencies as screens under extended illumination is that such an interpretation will fail seriously if the illuminant must first pass through the screen to reach B and A below. In this case, the filtered light reaching B and A will have a strong chromatic component from P+Q which with even minimal diffusion will seriously alter the spectral content of the light reflected back through P+Q. Because the spectral characteristics of P and B (or Q and A) are assumed selective and different, the reflected light will be reduced depending upon the extent of overlap of the two spectral reflectance functions $\rho_{PQ}(\lambda)$

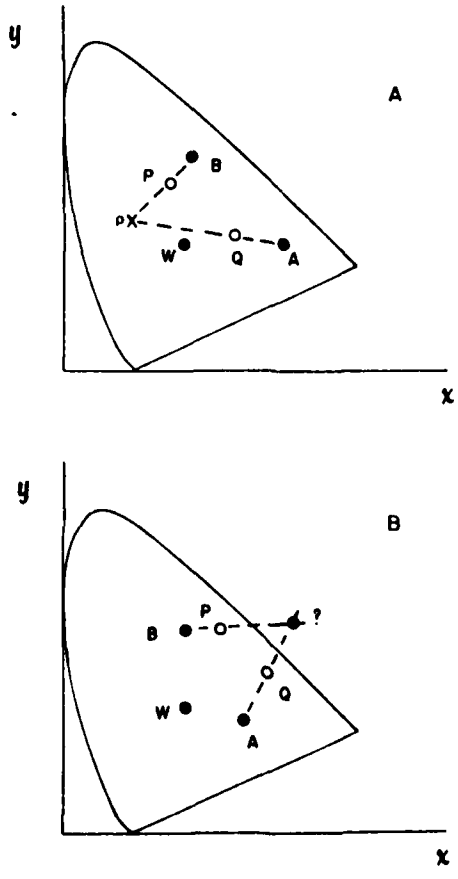


Figure 3. Upper figure shows the standard CIE spectrum locus in chromaticity coordinates x and y. According to classical color mixture laws, if P + Q is a transparent overlay above regions A and B, then the chromaticity of the reflective component of P + Q must lie at ρ. The lower figure shows that the additive mixture rule can lead to imaginary chromaticities.

and $\rho_B(\lambda)$ (or $\rho_A(\lambda)$ for A). The transmittance of P (and Q) will therefore always be underestimated and will appear to be zero when $\rho_P(\lambda)$ and $\rho_B(\lambda)$ do not overlap at all, for example.

The same argument applies even more forcefully if P•Q is a film (either transparent or translucent). In this case, the spectral intensity P(λ) in the image will always be less than the expected mixture sum calculated from an average transmittance τ and reflectance ρ:

$$P(\lambda) = E(\lambda)\cdot\tau(\lambda)\cdot\rho_B(\lambda) + E(\lambda)\cdot\rho(\lambda) < I_B(\lambda)\cdot\tau + E(\lambda)\cdot\rho \qquad (17)$$

Depending upon the overlap in the relevant spectral distributions, the mixture estimates of transmittance can be either too high or too low. For example, referring to equation (6) the difference (I_B-I_A) can be made small, but (I_P-I_Q) can be made very large by letting the spectral reflectance of ρ not overlap with that of A, but with considerable overlap with ρ_B of surface B. Finally, imaginary reflectances can result. Consider figure 3b, where two yellowish surfaces P•Q are superimposed upon a yellow-green surface B and an orange surface A. (As shown in the next section, this condition arises when a spectral filter of no reflectance is placed over A and B.) Observers have no trouble seeing such a configuration as transparent, yet the mixture rule predicts an imaginary spectral reflectance function that lies outside physically realizable surfaces (Bolt, et. al., 1977).

Additive color mixture rules are therefore inappropriate for evaluating the transmittance of screens or films, where subtactive effects must occur. Since Metelli's scission theory is essentially an additive mixture theory of transparency, it fails when spectral factors are considered.

5.2 Subtractive color mixtures

The introduction of a transparent film in front of a reflecting surface will always decrease the spectral bandwidth of the transmitted light, thereby increasing its purity. The farther apart the bandpass characteristics of the filter $\tau(\lambda)$ and the background reflectance $\rho_B(\lambda)$, the greater the increase in purity of the transmitted light and the less its magnitude. Under the common condition where the chromaticities of $\tau(\lambda)$ and $\rho_B(\lambda)$ lie on complementary sides of the "white" point, then the product of their bandpassed values $\tau_\lambda\cdot\rho_B(\lambda)$ will approach zero as purity increases. In the limiting case, then, the reflectance of ρ_λ of the film will be the sole determinant of the spectral content of P and Q. P will then equal Q, and P•Q must be considered opaque because no hint of the difference B-A has been passed.

In order for the film to pass some quality of B and A, the spectral transmittance τ_λ of the film must lie near or within the triangle WBA (figure 4A). (Quantitative specifications for ideal filters can be made, relating the maximum transmittance possible for any $\tau_\lambda \cdot \rho_B(\lambda)$ combination. The locus for this product exceeding 10% is not too distant from point W.) Now consider a film of zero reflectance and spectral transmittance τ_λ, with τ_λ lying within WBA. Depending upon the strength of τ_λ and the reflectances ρ_B and ρ_A for surfaces B and A, the imaged chromaticities of B and A (i.e., P(λ) and Q(λ)) must move from location τ_λ away from locations B and A to new positions B'=P and A'=Q. But the greater the movement toward increasing purity, then the less the net light transmitted. Saturation changes will thus always be correlated with the net change in image intensity. If this rule is violated, the region P+Q cannot be a film.

Note that in this case where the film has no reflectance and its transmittance lies within WBA, the backgrounds seen through the film must become more saturated, with the color difference decreasing (see figure 4A). This is just the configuration that leads to imaginary reflectances for additive mixtures, as illustrated in figure 3B. Thus, if imaginary reflectances are predicted by the relative positions of P,B and Q,A, the correct interpretation is of a film with zero reflectance, with the shift in chromatic positions being due solely to the transmittance of the film. Given the image intensities of P, Q, B and A, together with their chromaticities, it is possible (in principle) to solve for the location of τ assuming ideal filters.

Finally, it should be obvious by now that if the film has a reflective component ρ_λ, then it is not generally possible to find unique reflectances and transmittances of the transparency (Fig. 4B). Let B(λ) be the spectral reflectance of the background, $\rho(\lambda)$ and $\tau(\lambda)$ be the transmittance of the film, E(λ) the illuminant and $c_i(\lambda)$ the spectral sensitivity function of one of the color mechanisms (i=1,3) then the light transmitted from the background (assuming extended illumination) and detected by the i^{th} color mechanism is

$$T_{iB} = \int B(\lambda)\cdot\tau(\lambda)E(\lambda)\cdot c_i(\lambda)d\lambda$$

Similarly, the strength of the response to the reflected light is:

$$R_i = \int \rho(\lambda)\cdot E(\lambda)\cdot c_i(\lambda)d\lambda$$

In region P, the stimulus to the i^{th} color mechanism P_i then becomes:

$$P_i = D\cdot T_{iB} + (1-D)Ri, \quad i=1,3 \qquad (18)$$

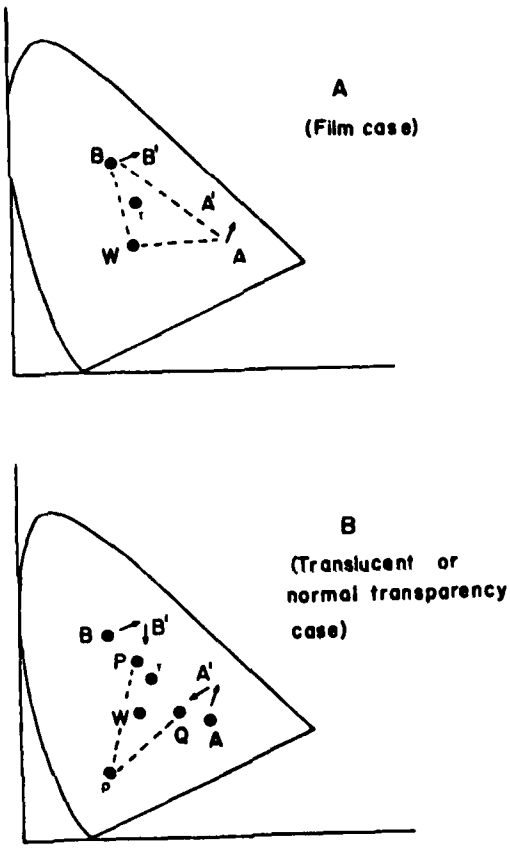


Figure 4. The effect of subtractive color mixture upon transparency is shown in the upper diagram (a). The interposition of a filter τ, must change the chromaticities of the backgrounds A and B to A^1 and B^1 as illustrated by the arrows. If the transparency has a reflective component, ρ, then additive mixture laws apply to ρ, B^1, and A^1 to constrain the color appearance of P and Q as shown in the lower diagram. The calculation of the positions of τ and ρ solely from P, Q and A, B is impossible in most cases.

which is analogous to equation (1), where D is the hole density. For two backgrounds A and B and two transparency regions P+Q we now have twelve stimulus strengths (scalar) altogether. The unknowns are five *functions* of λ: E(λ), B(λ), A(λ), $\rho(\lambda)$, $\tau(\lambda)$. Unless some constraints can be imposed upon the form of these functions, it is not generally possible to recover their values from only twelve scalars. (Beck's 1978 claim that additive and subtractive mixtures yield equivalent transparency solutions is incorrect because it ignores both λ and the illuminant.)

5.3 Imposing natural constraints

A. "Crack" constraint: In the achromatic case, the instance where region A was a "crack" was helpful. For color transparency, the "crack" imposes the constraint that

$$Q_i = (1-D)R_i$$

because the transmitted component is zero. Thus

$$P_i - Q_i = D \cdot T_{iB} = D \int B(\lambda) \cdot \tau(\lambda) \cdot E(\lambda) \cdot c_i(\lambda) d\lambda \qquad (19)$$

Now let τ_o be the achromatic transmittance (compute by taking the maximum value of $\tau(\lambda)$). Then $\tau(\lambda) \leq \tau_o$, and

$$T_i \leq \tau_o \cdot I_{iB} = \tau_o \int B(\lambda) E(\lambda) c_i(\lambda) d\lambda$$

and

$$P_i - Q_i = D \cdot T_{ib} \leq D \cdot \tau_o I_{iB} \quad i=1-3. \qquad (20)$$

where I_{iB} is the image intensity for region B seen by the i^{th} color channel. Since $D \cdot \tau_o$ is the average maximum transmittance τ_o, we can recast equation (20) in terms of chromatic contrast similar to the achromatic case (equation 12):

$$\tau_o \geq \tau_i = 2 \cdot \Delta C_{i,PQ} / (1 + \Delta C_{i,PQ}) \qquad (21)$$

where $\Delta C_{i,AB}$ is taken as 1, and a thin "film" is assumed. The change in chromatic contrast in any of the three color channels in the region of the "crack" can thus not exceed the maximum transmittance of the film. Note that this rules does not imply that the change in chromatic contrast

across P and Q must always be less than the change across B and A. For example, if B is greenish yellow and A is a reddish yellow of a non-overlapping spectral function, the interposing a blue-green filter can cause Q (over A) to become black while leaving P (over B) light.

B. Saturation Rule for Transparency: By extending equations (10)-(12) to the chromatic case, it can be shown that the change in chromatic contrast (in each color channel) due to transparency must be less than the change in contrast in the background (such as in the presence of a "crack"). This fact immediately constrains the relations between P and Q in the (idealized) chromaticity diagram. In Figure 5A, the location of the blue primary β is given and rays are drawn to background regions B and A. The crosshatched area then shows the maximum change in the blue channel ($\Delta\beta$) possible in the transparent region. (Ideally, logarithmic coordinates should be chosen.) In figure 5B, the maximum chromatic contrast for the red channel is illustrated. A similar rectangle can also be constructed for the green channel. Figure 5C shows the size of the limiting polygon. Its position can be moved within an approximate chromaticity diagram based on logarithmic coordinates.

C. Complementary Color Constraint: Imagine now that the allowable polygon for P+Q in figure 6A moved to the opposite side of "white" (W) as shown by the dashed outline in figure 6A. The colors are then complementary to those of the background, and no hint of the background colors is apparent. Although such a transparent medium is physically possible, transparency cannot be proven from image information alone. The more conservative procedure would thus be to require that transparent medium reveal at least a hint of the background colors A and B. This constraint would then limit the range of P+Q to colors lying on the same side of "white" as the background, as shown by the various positions of the polygon outline in figure 6B.

6. PRELIMINARY PSYCHOPHYSICS

6.1 Achromatic measurements

Using a "Texture Display" driven by a DEC PDP-11 [Richards, 1978], a pattern consisting of four regions similar to figure 1 was generated. To test for the general validity of equation (6A), all possible combinations of gray level arrangements were examined for transparency using four gray levels: .2, .4, .6, and .8 contrast. (This yielded 64 non-redundant arrangements.) Transparency was seen in all cases where equation (6A) was satisfied and was never seen when equation (6A) was not

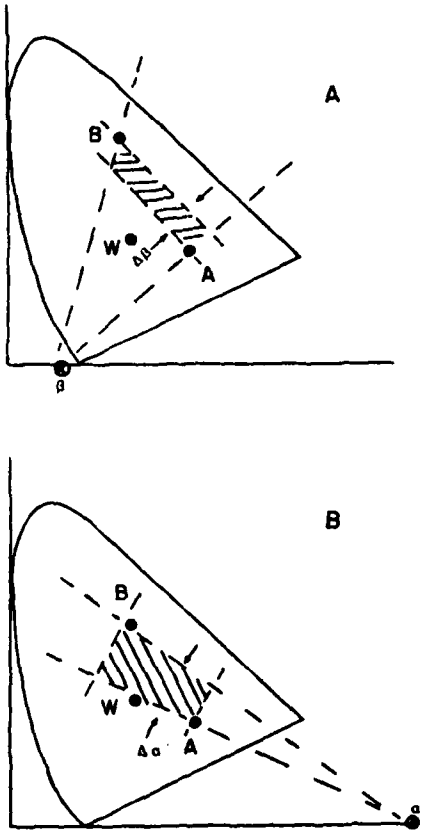


Figure 5. Saturation rule for transparency is illustrated in the upper diagram. The blue primary is located at the intersection of the dashed lines. The cross-hatched region shows schematically the maximum allowable change in the *relative* value of this primary due to interposing a transparency that acts to degrade contrast. In the lower diagram, maximum allowable change for the relative strength of the red primary is illustrated.

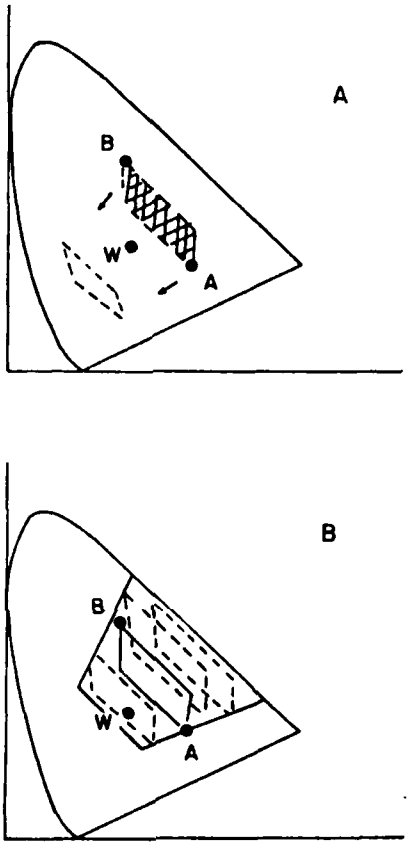


Figure 6. By combining the saturation constraints illustrated in Fig. 5, a polygon about the background chromaticities A and B can be constructed. The net result is illustrated in the upper diagram. In an ideal color coordinate system, the displacement of this polygon to any region sets a limitation on the distance between P and Q in any transparent overlay. In the lower diagram the possible positions of the polygon are limited to the same side of the white point, w, as the background A and B. This restriction is imposed by the rule that some hint of the color of both A and B must appear in P and Q, excluding P and Q values that are complementary to A and B.

satisfied.

Such measurements, however, do not test the hypothesis that contrast is used to assess transparency, because no subjective estimates of the transmittance were required. As a preliminary test for equation (11) and (16) thin neutral density (Wratten) filters were placed over four pairs of gray-white backgrounds of reflectance ranging from .07 to .90 with a contrast step of approximately 0.2. Figure 7 shows the transmittance estimates for two observers averaged over the four background conditions. Note that estimated transmittance for most of the range follows a power relation with an exponent of about 1/1.5. This solution is intermediate to the case where the illumination is extended (predicted slope is 1.0) and where the illuminant must pass through the filter twice--once to illuminate the background and once again to be seen by the viewer (predicted slope is 1/2). The human observer's estimate is thus a compromise between these two common cases; the result is most likely due simply to the compressive transformation of image intensity into a lightness scale. In other words, if transmittance is estimated using apparent contrast, a result like Fig. 7 would be expected.

6.2 Chromatic measurements

Using an Apple II computer, various combinations of colors were painted in the regions P, Q, B, and A of figure 1. Although only 15 colors were available for any one setting of the TV color control knob, by adjusting the tints over the full range, a wide assortment of color combinations could be examined. The principal findings were:

1) Transparency was never seen if the overlay (P+Q) did not contain at least some hint of the background colors (B+A). (This is the complementary color constraint introduced in section 5.3c.)

2) The overlay (P+Q) also was required to have a common color for transparency to be seen.

3) The intensity relations should roughly (but not exactly) satisfy equation (6A).

These preliminary findings regarding the spectral factors used by man in assessing transparency verify the major set of constraints upon the positions of P+Q in the chromaticity diagrams outlined in the earlier sections. However, there appears to be a considerable relaxation in the allowable positions of P and Q in the rectangular region defined by B, A, and W for P+Q (see figure 8).

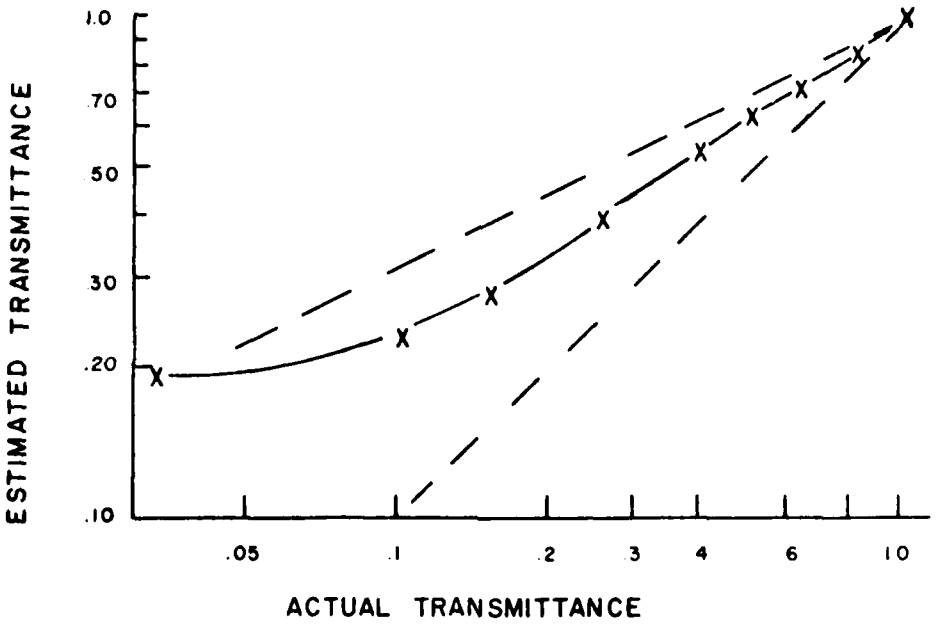


Figure 7. Transmittance estimates for two observers (WR, AW) for this films having essentially no reflective component. The data fall intermediate between the assumption of extended illumination (lower dashed line) and the illumination of the background by light that must first pass through the filter (upper dashed line). In the actual experiment, the illumination was extended.

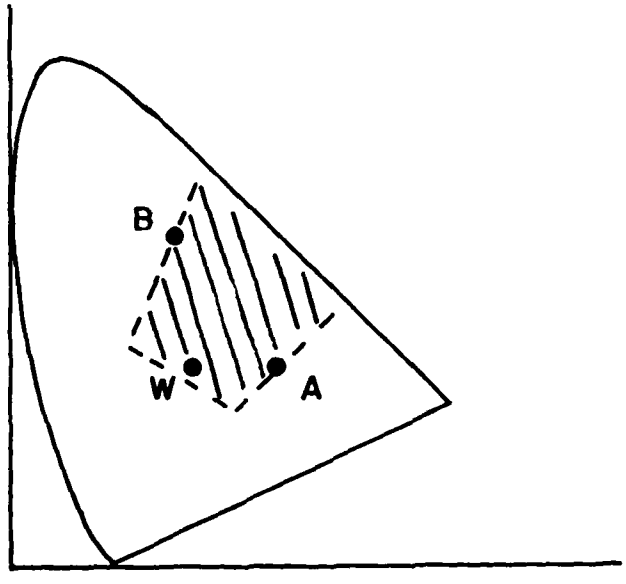


Figure 8. The cross-hatched region shows the empirical observation of the range of allowable chromaticities for the appearance of transparency given the two background chromaticities A and B. To appear transparent, the overlay must reveal some hint of the underlying background colors.

Perhaps by a closer scrutiny of the relations between the achromatic and chromatic factors in both transparency and translucency, the reason for the relaxation of the constraints can be determined.

6.3 Dependence of perceived transparency on local processes

The psychophysical investigation of perceptual transparency revealed an unexpected local component in the human observer's transparency computation. As discussed earlier in section 2.2, two regions P and B do not determine a unique transparency solution. Four regions P,Q,A,B are sufficient to determine either a unique transmittance and relative reflectance, if a solution exists, or to determine that P+Q cannot be a homogeneous transparent medium. If two additional regions, P' and Q' are added, the equations are over-determined. It is therefore possible to arrange the intensities of the six regions so that any combination of four (P-Q-A-B, P'-Q'-A-B, or A-Q Q'-B) determine a unique solution, but all the regions taken together are inconsistent.

Surprisingly, the human observer is entirely insensitive to such inconsistency, unless the inconsistency can be discovered by examining some single region of the image subtending less than about one degree. For example, in Fig. 2B set the image intensities of P=P' but let Q and Q^1 differ. Then A cannot be physically transparent, yet it will still appear as a homogeneous transparent filter unless local evidence of the inconsistency is available. Only the spatial relations separating P and P' need to be changed to alter the perception of transparency. When S is large, then local simultaneous comparison of all six regions is not possible, and transparency is perceived. When the regions are moved closer, the inconsistency is detected, and perceptual transparency is lost. This effect demonstrates that perceived transparency depends at least in part on a process of local comparison of intensities.

7. SUMMARY

With uncertainty in the exact illuminant conditions, the most useful cue for assessing the transparency of a medium is contrast reduction--either achromatic or chromatic. For many instances of transparency or translucency (haze, fog, or rain) the maximum contrast in the scene is a useful measure of the transmittance.

Spectral information can provide further support for achromatic deductions of transparency.

Man appears to use spectral information conservatively, however, and his rules exclude many instances where transparency is possible. The principal rules for transparency are that 1) a portion of the background color must "leak through" and 2) the transparent medium must possess at least a hint of the one color throughout the region of the transparent overlay. These rules favor the assessment of "whitish" transparent media such as haze, fog or rain.

Acknowledgements: Dr. G. Haddad and Major J. Thorpe provided helpful criticism on drafts of this report. The technical assistance of C. Papineau was also greatly appreciated.

REFERENCES

Beck, J. 1978 Additive and subtractive mixture in color transparency. *Percept. & Psychophysics 23*, 265-267.

Bolt, R. A., Negroponte, N. and Tom, V. 1977 Color transparency effects from mosaics on opaque color. *Final Report, U. S. Army Contract #DAAG29-76-0037.* M.I.T. Architecture Machine Group.

Gibson, J.J. 1950 *The perception of the visual world.* Boston: Houghton Mifflin.

Hilbert D., & Cohn-Vossen, S. 1952 *Geometry and the Imagination.* Chelsea Publishing.

Horn, B.K.P. 1975 Obtaining shape from shading information. In *The psychology of computer vision*, P. H. Winston, ed. New York: McGraw-Hill.

Kanade, T., and Kender, J.R. 1979 Skewed symmetry: Mapping image regularities into shape, Technical Report, Computer Science Department, Carnegie-Mellon University, (forthcoming).

Koffka, K. 1935 *Principles of Gestalt Psychology.* New York: Harcourt, Brace and World.

Kennedy, J.M. 1974 *A psychology of picture perception.* San Francisco: Jossey-Bass.

Kubella, P. & Munk, F. 1934 Ein Bertiag zur Optik des Farbanstriche. *Z. techn. Physik 12*, 593.

Kubella, P. 1954 New contributions to the optics of intensely light-scattering materials, Part II. Non-homogeneous layers. *J. Opt. Soc. Am. 44*, 330.

Land, E. H. & McCann, J. J. 1971 Lightness and retinex theory. *J. Opt. Soc. Am. 61*, 1-11.

Marr, D. 1976 Artificial Intelligence - a personal view. *Artificial Intelligence 9*, 37-48. Also available as M.I.T. A.I. Memo 355.

Marr, D. 1976 Early processsing of visual information. *Phil. Trans. Roy. Soc. B. 275*, 483-524.

Marr, D. 1976 Analysis of occluding contour. *Proc. R. Soc. Lond. B. 197*, 441-475. Also available as M.I.T. A.I. Lab. Memo 372.

Marr, D. & Poggio, T. 1977 From understanding computation to understanding neural circuitry. *Neurosci. Res. Prog. Bull. 15*, 470-488.

Metelli, F. 1970 An algebraic development of the theory of transparency. *Ergonomics 13 (1)*, 59-66.

Metelli, F. 1974 The perception of transparency. *Sci. Am. 230 (4)*, 125-135.

Postman, L. & Tolman, E.C. 1959 Brunswick's Probabilistic Functionalism. In *Psychology: A study of a science.* Edited by S. Koch. New York: McGraw-Hill.

Richards, W. 1979 Why rods and cones? *Biol. Cyber. 33*, 125-135.

Richards, W. 1978 Experiments in texture perception. *Final Report, AFOSR Contract F44620-74-C-0076*, pp. 1-116. (Available from M.I.T. Psychology Department, Cambridge, Mass. 02139.)

Stevens, K.A. 1978 Computation of locally parallel structure. *Biological Cybernetics 29*, 19-28. Also available as M.I.T. A.I. Lab Memo 392.

Rock, I. 1975 An Introduction to Perception. MacMillan:N.Y.

Stevens, K.A. 1979 Surface perception from local analysis of texture and contour. M.I.T. A.I. Lab TR-512.

Waltz, D. 1975 Understanding line drawings of scenes with shadows. In *The Psychology of computer vision*, P.H. Winston, ed. New York: McGraw-Hill.

Witkin, A. 1979 Shape from contour. M.I.T. Ph.D. Thesis (Psychology), forthcoming.

Wysecki, G. & Stiles, W. S. 1967 *Color Science*, New York: Wiley.